

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

STIC-Biotech/ChemLib

100 291

From: Nickol, Gary
Sent: Friday, August 01, 2003 11:24 AM
To: STIC-Biotech/ChemLib
Subject: 09/988975

RECEIVED

AUG - 1 2003

Please search and interference search the following:

(STIC)

SEQ ID NO:1

Thanks,

Gary Nickol
CM1, AU:1642
Room 8D17, Mailbox 8E12
(703) 305-7143

Point of Contact
P. Sheppard
Telephone number: (703) 308-4499

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 8/4/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:10:54 ; Search time 19 Seconds

(without alignments)
630.209 Million cell updates/sec

Title: US-09-988-975a-1

Perfect score: 1543

Sequence: 1 MKLNTSAGVNDIYKKLYHSD.....YKGVNVPFTDWTYKQWRADG 283

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:*

- 1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/prodata/1/1aa/PCTUS.COMB.pep:*
- 6: /cgn2_6/prodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1541	99.9	283	3	US-08-807-151-1
2	1541	99.9	283	4	US-09-478-957-1
3	1540	99.8	492	3	US-09-342-749-2
4	1540	99.8	492	4	US-09-691-840-2
5	857	55.5	159	3	US-09-518-046-24
6	696	45.1	454	3	US-09-518-046-2
7	588.5	38.1	455	3	US-09-261-416-2
8	558.5	36.2	423	4	US-09-656-002-2
9	558.5	36.2	435	3	US-09-008-271A-6
10	544.5	35.3	288	3	US-08-844-483-63
11	541	35.1	638	2	US-08-681-151-3
12	536.5	34.8	798	1	US-08-200-900A-2
13	536.5	34.8	798	5	PCT-US94-00616-2
14	522.5	33.7	376	4	US-09-820-002-4
15	520.5	33.7	417	4	US-09-820-002-4
16	519.5	33.7	256	2	US-09-027-337-3
17	519.5	33.7	256	4	US-09-644-600-3
18	518.5	33.6	488	1	US-08-508-448C-55
19	518.5	33.6	418	4	US-09-370-838-62
20	518.5	33.6	418	4	US-09-370-838-82
21	518.5	33.6	418	4	US-09-370-838-83
22	515	33.4	288	3	US-08-944-483-64
23	514.5	33.3	255	3	US-08-944-483-67
24	512.5	33.2	235	3	US-08-807-151-3
25	512.5	33.2	225	4	US-09-478-957-3
26	512.5	33.2	855	4	US-09-027-337-2
27	512.5	33.2	855	4	US-09-644-600-2

28	505.5	32.8	812	1	US-08-248-629A-1	Sequence 1, Appl
29	505.5	32.8	812	1	US-08-451-932-1	Sequence 1, Appl
30	505.5	32.8	812	1	US-08-452-260-1	Sequence 1, Appl
31	505.5	32.8	812	1	US-08-326-785-1	Sequence 1, Appl
32	505.5	32.8	812	2	US-08-612-788-1	Sequence 1, Appl
33	505.5	32.8	812	2	US-08-605-598B-1	Sequence 1, Appl
34	505.5	32.8	812	2	US-08-429-743-1	Sequence 1, Appl
35	505.5	32.8	812	2	US-08-866-738-1	Sequence 1, Appl
36	505.5	32.8	812	3	US-09-066-028-1	Sequence 1, Appl
37	505.5	32.8	812	4	US-09-192-012-3	Sequence 3, Appl
38	505.5	32.8	812	4	US-09-335-325-1	Sequence 12, Appl
39	505.5	32.8	812	4	US-08-991-761A-12	Sequence 1, Appl
40	505.5	32.8	812	5	PCT-US95-05107-1	Sequence 19, Appl
41	504.5	32.7	232	1	US-08-508-448C-19	Sequence 65, Appl
42	504.5	32.6	235	3	US-08-944-483-65	Sequence 2, Appl
43	501.5	32.5	416	2	US-09-000-846-2	Sequence 10, Appl
44	498.5	32.3	902	4	US-09-644-600-10	Sequence 26, Appl
45	491.5	31.9	285	4	US-09-023-942A-26	

ALIGNMENTS

RESULT 1
US-08-807-151-1
; Sequence 1, Application US/08807151
; Patent No. 604303
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,151
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0227 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNIOTOL
; CLONE: 556016
; US-08-807-151-1
Query Match 99.9%, Score 1541, DB 3, Length 283;
Best Local Similarity 100.0%, Pred. No. 8.5e-161;
Matches 283, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1 MKLNTSAGNVDIYKCLYHSDACSSRAVVSRLRCIACGVNLNSRSRIVGGSALPGAMPW 60
DB 1 MKLNTSAGNVDIYKCLYHSDACSSRAVVSRLRCIACGVNLNSRSRIVGGSALPGAMPW 60
QY 61 QVSLHVNQVHVGCGSITTEPMTVTAHCEKPLNNPMTAFAGILROSFPYAGAYOVE 120
DB 61 QVSLHVNQVHVGCGSITTEPMTVTAHCEKPLNNPMTAFAGILROSFPYAGAYOVE 120
QY 121 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPYCLPMPGMLQPEOLCWSGMAITEE 180
DB 121 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPYCLPMPGMLQPEOLCWSGMAITEE 180
QY 181 KGTSEVLAARAVLLETCRCNSRYVNDLITPAMTCAGFLQGVNDSQQDGGGLVTSK 240
DB 181 KGTSEVLAARAVLLETCRCNSRYVNDLITPAMTCAGFLQGVNDSQQDGGGLVTSK 240
QY 241 NNIMWLIGDTSWGGCAKAYRPGVYGVNVVFTDWTYRQMRADG 283
DB 241 NNIMWLIGDTSWGGCAKAYRPGVYGVNVVFTDWTYRQMRADG 283

RESULT 2

US-09-478-957-1
Sequence 1, Application US/09478957
Patent No. 6350448
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/478,957
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,151
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0227 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SCORNO101
CLONE: 596016
US-09-478-957-1

Query Match 99.9%; Score 1541; DB 4; Length 283;
Best Local Similarity 100.0%; Pred. No. 8,5e-161; Indels 0; Gaps 0;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKCLYHSDACSSRAVVSRLRCIACGVNLNSRSRIVGGSALPGAMPW 60
DB 1 MKLNTSAGNVDIYKCLYHSDACSSRAVVSRLRCIACGVNLNSRSRIVGGSALPGAMPW 60
QY 61 QVSLHVNQVHVGCGSITTEPMTVTAHCEKPLNNPMTAFAGILROSFPYAGAYOVE 120
DB 61 QVSLHVNQVHVGCGSITTEPMTVTAHCEKPLNNPMTAFAGILROSFPYAGAYOVE 120
QY 121 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPYCLPMPGMLQPEOLCWSGMAITEE 180
DB 121 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPYCLPMPGMLQPEOLCWSGMAITEE 180
QY 181 KGTSEVLAARAVLLETCRCNSRYVNDLITPAMTCAGFLQGVNDSQQDGGGLVTSK 240
DB 181 KGTSEVLAARAVLLETCRCNSRYVNDLITPAMTCAGFLQGVNDSQQDGGGLVTSK 240
QY 241 NNIMWLIGDTSWGGCAKAYRPGVYGVNVVFTDWTYRQMRADG 283
DB 241 NNIMWLIGDTSWGGCAKAYRPGVYGVNVVFTDWTYRQMRADG 283

RESULT 3

US-09-342-749-2
Sequence 2, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Teng, David H.-P.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: THERS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-342-749-2

Query Match 99.8%; Score 1540; DB 3; Length 492;
Best Local Similarity 99.6%; Pred. No. 2.5e-160; Indels 0; Gaps 0;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKCLYHSDACSSRAVVSRLRCIACGVNLNSRSRIVGGSALPGAMPW 60
DB 210 MKLNTSAGNVDIYKCLYHSDACSSRAVVSRLRCIACGVNLNSRSRIVGGSALPGAMPW 269
QY 61 QVSLHVNQVHVGCGSITTEPMTVTAHCEKPLNNPMTAFAGILROSFPYAGAYOVE 120
DB 270 QVSLHVNQVHVGCGSITTEPMTVTAHCEKPLNNPMTAFAGILROSFPYAGAYOVE 329
QY 121 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPYCLPMPGMLQPEOLCWSGMAITEE 180
DB 121 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPYCLPMPGMLQPEOLCWSGMAITEE 369
QY 181 KGTSEVLAARAVLLETCRCNSRYVNDLITPAMTCAGFLQGVNDSQQDGGGLVTSK 240
DB 181 KGTSEVLAARAVLLETCRCNSRYVNDLITPAMTCAGFLQGVNDSQQDGGGLVTSK 449
QY 241 NNIMWLIGDTSWGGCAKAYRPGVYGVNVVFTDWTYRQMRADG 283
DB 450 NNIMWLIGDTSWGGCAKAYRPGVYGVNVVFTDWTYRQMRADG 492

RESULT 4

US-09-691-840-2
Sequence 2, Application US/09691840
Patent No. 6444419

GENERAL INFORMATION:
 APPLICANT: Wong, Alexander K.C.
 APPLICANT: Ravitskian, Sean V.
 APPLICANT: Teng, David H.-F.
 APPLICANT: Myriad Genetics, Inc.
 TITLE OF INVENTION: Tmprss2 is a Tumor Suppressor
 FILE REFERENCE: 2318-202
 CURRENT APPLICATION NUMBER: US/09/691,840
 CURRENT FILING DATE: 2000-10-18
 PRIOR APPLICATION NUMBER: US/09/342,749
 PRIOR FILING DATE: 1999-06-29
 PRIOR APPLICATION NUMBER: US 60/091,044
 PRIOR FILING DATE: 1998-06-29
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 492
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-691-840-2

Query Match 99.8%; Score 1540; DB 4; Length 492;
 Best Local Similarity 99.6%; Pred. No. 2, 5e-160;
 Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNINSRQRIYVGSALPGAMPW 60
 DB 210 MKLNTSAGNDVIYKLYHSDACSSKAVSLRCLACGVNINSRQRIYVGSALPGAMPW 269
 QY 61 QVSLHAYQNVHVCSSITTEPMITVTAHCVEKPLNPMHTAFAGILRQSPFVYAGVYE 120
 DB 270 QVSLHAYQNVHVCSSITTEPMITVTAHCVEKPLNPMHTAFAGILRQSPFVYAGVYE 329
 QY 121 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPVCLPMPGMLQPBOLCWSMGATEE 180
 DB 330 KVISHPYDSKTNNDIALMKLOKPLTFNDLVKPVCLPMPGMLQPBOLCWSMGATEE 389
 QY 181 KGTSEVLANAKYLLLETORCNSRYVDLITPAMICAGFLQGNVDSQSGSGGLVTSK 240
 DB 390 KGTSEVLANAKYLLLETORCNSRYVDLITPAMICAGFLQGNVDSQSGSGGLVTSK 449
 QY 241 NNTWMLIGDTSMGSCAKAYRPGVGNVVFDTWIRQWRADG 283
 DB 450 NNTWMLIGDTSMGSCAKAYRPGVGNVVFDTWIRQWRADG 492

RESULT 5
 US-09-518-046-24
 Sequence 24, Application US/09518046
 Patent No. 6294663
 GENERAL INFORMATION:
 APPLICANT: O'Brien, Timothy J.
 APPLICANT: Underwood, Lowell J.
 TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
 FILE REFERENCE: D6192CIP
 CURRENT APPLICATION NUMBER: US/09/518,046
 CURRENT FILING DATE: 2000-03-02
 EARLIER APPLICATION NUMBER: 09/261,416
 EARLIER FILING DATE: 1999-03-03
 NUMBER OF SEQ ID NOS: 153
 SEQ ID NO 24
 LENGTH: 159
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: DOMAIN
 OTHER INFORMATION: protease domain of Tmprss2 (Tmprss2)
 US-09-518-046-24

Query Match 55.5%; Score 857; DB 3; Length 159;
 Best Local Similarity 98.7%; Pred. No. 4, 2e-86;
 Matches 157; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 80 EMITVAHCVCKPLNNPMTAFAGILRQSPFVYAGVYVEKVISHPYDSKTNNDIAL 139
 DB 1 EMITVAHCVCKPLNNPMTAFAGILRQSPFVYAGVYVEKVISHPYDSKTNNDIAL 60
 QY 140 MKLOKPLTFNDLVKPVCLPMPGMLQPBOLCWSMGATEESGKTSVLANAKYLLIETO 199
 DB 61 MKLOKPLTFNDLVKPVCLPMPGMLQPBOLCWSMGATEESGKTSVLANAKYLLIETO 120
 QY 200 RNSRYYVDNLITPAMICAGFLQGNVDSQSGSGGLVTSK 238
 DB 121 RNSRYYVDNLITPAMICAGFLQGNVDSQSGSGGLVTSK 159

RESULT 6
 US-09-518-046-2
 Sequence 2, Application US/09518046
 Patent No. 6294663
 GENERAL INFORMATION:
 APPLICANT: O'Brien, Timothy J.
 APPLICANT: Underwood, Lowell J.
 TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
 FILE REFERENCE: D6192CIP
 CURRENT APPLICATION NUMBER: US/09/518,046
 CURRENT FILING DATE: 2000-03-02
 EARLIER APPLICATION NUMBER: 09/261,416
 EARLIER FILING DATE: 1999-03-03
 NUMBER OF SEQ ID NOS: 153
 SEQ ID NO 2
 LENGTH: 454
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: complete amino acid sequence of TADG-12
 OTHER INFORMATION: protein
 US-09-518-046-2

Query Match 45.1%; Score 696; DB 3; Length 454;
 Best Local Similarity 50.4%; Pred. No. 9e-68;
 Matches 137; Conservative 37; Mismatches 92; Indels 6; Gaps 4;

QY 12 IYKLYHSDACSSKAVSLRCLACGVNINSRQRIYVGSALPGAMPQVSLHAYQNVH 71
 DB 185 LHSVYVRBGCAHGVYTLQCTACGRRGYS--SLIVGMSLSLQWYKSLQFQGHIL 242
 QY 72 CGSITTEPMITVTAHCVEKPLNPMHTAFAGILRQSPFVYAGVYVEKVISHPYDS 130
 DB 243 CGSITTEPMITVTAHCVEKPLNPMHTAFAGILRQSPFVYAGVYVEKVISHPYDS 299
 QY 131 KTKNNDIALMKLOKPLTFNDLVKPVCLPMPGMLQPBOLCWSMGATEESGKTSVLANA 190
 DB 300 KTKNNDIALMKLOKPLTFNDLVKPVCLPMPGMLQPBOLCWSMGATEESGKTSVLANA 359
 QY 191 AKVLLITFORCNSRYVDNLITPAMICAGFLQGNVDSQSGSGGLVTSKNNIWMILGDT 250
 DB 360 AAVLLISNKCINRHDYVGGIISPSMLCAGYLVGVDSCQSGSGGLVQGBRRLWLVGAT 419
 QY 251 SMGSGCAKAYRPGVGNVVFDTWIRQWRADG 282
 DB 420 SMGSGCAKAYRPGVGNVVFDTWIRQWRADG 451

RESULT 7
 US-09-261-416-2
 Sequence 2, Application US/09261416A
 Patent No. 6291663
 GENERAL INFORMATION:
 APPLICANT: O'Brien, Timothy J.
 APPLICANT: Underwood, Lowell J.
 TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
 TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
 FILE REFERENCE: D6192

CURRENT APPLICATION NUMBER: US/09/261,416A
CURRENT FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 14
SEQ ID NO 2
LENGTH: 455
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Amino acid sequence of TADG-12 encoded by nucleotides
Patent No. 6291663
US-09-261-416-2

Query Match 38.1%; Score 588.5; DB 3; Length 455;
Best Local Similarity 46.2%; Pred. No. 5.7e-56;
Matches 127; Conservative 35; Mismatches 102; Indels 11; Gaps 7;

QY 12 IYKLVHSDACSKAVVSLRCLACGVNLNNSRSGRIYVGSBALPGAMPVQVSLHVNQVH 71
DB 185 LHSVYVREGCAGHVVTLOCTACGHRGYS--SRIVGNSMISLQWPMQASLQFGVHL 242
QY 72 CGGSITPEVITAAHCEKPLNPNHMTAFAGILROSEMTGA-GYQVENVISHPNYDS 130
DB 243 CGGSVITPEVITAAHCV-YDYLPRKMTIQVGLV--SLDNPASHLVEKIVHRSKXP 299
QY 131 KTRNDIALMKLOKPLTNDLVKPVCLPNPQWMLQPEOLCWSIGWATEEKGKTSVTLNA 190
DB 300 KRLGNDIALMKLAGPLTFNEMIQVCLPFSSENFDDKVCWTSIGWATEDEGDSAPVTLN 359
QY 191 AKVLIETQ--RCKSVYVDNLITPMTICAGFLQG--NDSGCGSGGLVTSKNNIMWLI 247
DB 360 AAVPLLSNKLQOGQVRRMHL--PLHALRGLPDCMRWNSCGSGGGLVQGBERRLWKL 417
QY 248 GDTSGSGCAKAYRPGVGNVVFDTMIRQNRAD 282
DB 418 GATSPGIGCADVVKPGVYTRVTSFLMIRHQBMRD 452

RESULT 8
US-09-656-002-2
Sequence 2, Application US/09656002
Patent No. 6455668
GENERAL INFORMATION:
APPLICANT: Mack, David
APPLICANT: Giesh, Kurt
APPLICANT: Wilson, Keith
TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSING COLORECTAL CANCER, COMPOSITIONS, AND
FILE REFERENCE: A-69108/DJB/JTD/AMS
CURRENT APPLICATION NUMBER: US/09/656,002
CURRENT FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/525,993
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: US 09/493,444
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: PCT/US 00/07044
PRIOR FILING DATE: 2000-03-15
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 423
TYPE: PRT
ORGANISM: Homo sapiens
US-09-656-002-2

Query Match 36.2%; Score 558.5; DB 4; Length 423;
Best Local Similarity 43.9%; Pred. No. 1e-52;
Matches 118; Conservative 41; Mismatches 93; Indels 17; Gaps 8;
QY 19 SDACSSKAVVSLRCLACGVNLNNSRSGRIYVGSBALPGAMPVQVSLHVNQVHVGGSIT 78
DB 166 SGCLSGSLVSLHCLAGKSL--KTPRVGSGEASVDSWPMQVSIQYDKHGVGGSITD 222

QY 79 PEMVITAAHCEKPLNPNHMTAFAGILR-OSFMEYQAGYQVKEYI---SHPNYDSKTKN 134
DB 223 PHVITAAHCEKRLH-IDVFWKVRAGSDLGSPF-----PSLAVAKIILIFRPNV---PKD 274
QY 135 NDIALMKLOKPLTNDLVKPVCLPNPQWMLQPEOLCWSIGWATEEKGKTSVTLNAKY 193
DB 275 NDIALMKLOKPLTFSGVNPLCLPFDBELTPATPLWIIIGWFTKONGKNSDILLONSV 334
QY 194 LLIFTRCNSRYVDNLITPMTICAGFLQGNVDSGCGSGGLVTSKNNIMWLIQDTSWG 253
DB 335 QVIDSTRCNADAVQGEVTEKXKWCAGIPGAGVDTCCGDSGGPLMYQSDQ--WHVYGVISWG 393
QY 254 SGCAKAYRPGVGNVVFDTMIRQNRAD 282
DB 394 YGCGGSPTPGVYTKVSATYLMWLYNVRAS 422

RESULT 9
US-09-008-271A-6
Sequence 6, Application US/09008271A
Patent No. 6203979
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Yue, Henry
Hillman, Jennifer L.
Guegler, Karl J.
Corley, Neil C.
Tang, Tom Y.
Shan, Pu-Yi
TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSER: Incycle Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,271A
FILING DATE: 16-Jan-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mohan-Petersen, Sheila
REGISTRATION NUMBER: 41,201
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLN00T13
CLONE: 1337018
SEQUENCE DESCRIPTION: SEQ ID NO: 6
US-09-008-271A-6

Query Match 36.2%; Score 558.5; DB 3; Length 435;
Best Local Similarity 43.9%; Pred. No. 1e-52;
Matches 118; Conservative 41; Mismatches 93; Indels 17; Gaps 8;
QY 19 SDACSSKAVVSLRCLACGVNLNNSRSGRIYVGSBALPGAMPVQVSLHVNQVHVGGSIT 78

Db 178 SGPCLSSLSVSLHAGSGL---KTPRVVGGSEASVDSWPMQVSIQYKXKOVCGSLTD 234
Qy 79 PBNVLTAAHCYCKELNPNMTAPAGILR-QSMFYAGYOYKVI---SHPNYSKTKN 134
Db 235 PBNVLTAAHCYCKELNPNMTAPAGILR-QSMFYAGYOYKVI---SHPNYSKTKN 134
Qy 135 NDIALMLOKPLTENDLVKPVCLPNPQMLQPEQLCMISMGATSEK-GKTSVYANAKV 193
Db 287 NDIALMLOKPLTENDLVKPVCLPNPQMLQPEQLCMISMGATSEK-GKTSVYANAKV 193
Qy 194 LLIETQCNRSRYVDNLIITPAMICAGFLQGNVDSGQSGGKLVTSKNNIMLIGDITSWG 253
Db 347 QVIDSTCNMDAAGQSVTERKMMKAGIPGQGVDTQGDGGLPAMQSDQ-MHVVIVSAG 405
Qy 254 SGCAKAVRPGVYGVNWFETDNIYRQMRAD 282
Db 406 YGCGSPSTPGVYTVSAVYIMVIMWKAE 434

RESULT 10

US-08-944-483-63
Sequence 63, Application US/08944483
Patent No. 6232456

GENERAL INFORMATION:

APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: KLAS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESS: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PatsSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483

FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:

INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6232456
US-08-944-483-63

Query Match 35.3%, Score 544.5, DB 3, Length 248;
Best Local Similarity 42.8%, Pred. No. 1,6e-51;
Matches 104; Conservative 46; Mismatches 84; Indels 9; Gaps 6;
Qy 47 IYCGSALPGAMPQVSLNVQ---NVAVCGSITPBNVLTAAHCYK-PLNPNMTAP 102
Db 1 IYCGTSSSGMGMQVSLQVLTQRLCGSLIGHMVLTAAHCYDGLPDDVNR--LY 58
Qy 103 AGIAGSMFYAGY-QYKVISHPNYSKTKNDIALMLOKPLTENDLVKPVCLPNP 161
Db 59 SGIIMLSDITDTPFSQIKELIHONKYVBEGNDIALITLQAPLVYTFQKPICLPSHG 118
Qy 162 NMLQPEQLCMISMGATSEKGTSEVYANAKVLLIETQCNRSRYVDNLIITPAMICAGFL 221
Db 119 DSTIITNCWYTGMPGSKRKEIOTNLQKNIPLVYNSQKRY-ODYKLTQRMVCAGTK 177
Qy 222 QGNVDSGQSGGKLVTSKNNIMLIGDITSWGCAKAVRPGVYGVNWFETDNIYRQMR- 280
Db 178 EGGKDACRKGDSGGLPVCKHNGMRLVGIITSGEGCARREQGVYTVSAVYIMVIMWKAE 237
Qy 281 ADG 283
Db 238 SDG 240

RESULT 11

US-08-681-151-3
Sequence 3, Application US/08681151
Patent No. 5869637

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Braxton, Scott Michael
APPLICANT: Goll, Surya
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PatsSeq Version 1.5

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,151

FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0074US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 205011


```

; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: Difrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-2

```

```

Query Match      33.9%; Score 522.5; DB 4; Length 376;
Best Local Similarity 37.4%; Pred. No. 7.5e-49;
Matches 108; Conservative 44; Mismatches 106; Indels 31; Gaps 5;

```

```

QY 22 CSSKA---VSLRACIAGYNNLSR-----QSRVGGESALPGAMPQ 61
DB 77 CSSRSNARVAGLCEEMGLSDCPGRFLAICODGRRKLPVDRIVGSDTSIGRPMQ 136
QY 62 VSLHVNIVHVGSGSIITPEMIVTAHCVKPLNPMHTAFAGILROSFMFYGAGYQVEK 121
DB 137 VSLRYGGAHLGGSLGSDMVLTAHCFPERNRVLSRWRVFAGAVQAQSP-HGLQGVQA 195
QY 122 VISHPNY-----DSKTKNDIALMKLOKELTFNDLVKVCCLPNPGMLOPEQLOMISGW 175
DB 196 VVYHGGYLPKPDENSEENNDIALVHLSPLPTRYQVCLPAAGALVDGKICTVYTGW 255
QY 176 GATEKGTSEVLNAKVLIIETORCNSRYVDNLITPAMICAGFLQGNVDSGQDSGGX 235
DB 256 GNTQYGGQAGVLOEARVPILSNVCGNADFYGNQIKPMFCAGYPEGGIDACQDSGGP 315
QY 236 LV-----TSKNITMWLIGDTSWGSGCAKARPGVGNMVFPTDWYRQMR 280
DB 316 FVCEDSISRTPRKRLCGIVSWGICLAKQKPGVYTKVSDPREMIPOAIK 364

```

```

RESULT 15
US-09-820-002-4
; Sequence 4, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: Difrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 417
; TYPE: PRT
; ORGANISM: HUMAN
US-09-820-002-4

```

```

Query Match      33.7%; Score 520.5; DB 4; Length 417;
Best Local Similarity 38.7%; Pred. No. 1.4e-48;
Matches 104; Conservative 43; Mismatches 109; Indels 13; Gaps 4;

```

```

QY 22 CSSKAVVSLRACIAGYNNLSRSGRIYGGESALPGAMPQVSLHVNIVHVGSGSIITPEV 81

```

```

DB 140 CPERGFLAICODCG--RRLPVDRIYGGSDTSIGRPMQVSLRFDGANTCGGSLSGDW 197
QY 82 IVTAHCVKPLNPMHTAFAGILROSFMFYGAGYQVEKVISHPNY-----DSKTKN 135
DB 198 VTLAAHCFPERNRVLSRWRVFAGAVQAQSP-HGLQGVQAVVYHGGYLPKPDENSEEN 256
QY 136 DIALMKLOKELTFNDLVKVCCLPNPGMLOPEQLOMISGWTGKTSBYLNAKVL 195
DB 257 DIALVHLSPLPTRYQVCLPAAGALVDGKICTVYTGWNTQYGGQAGVLOEARVPI 316
QY 196 IETORCNSRYVDNLITPAMICAGFLQGNVDSGQDSGGXLY-----TSKNITMWLIGDTS 251
DB 317 ISNDVCGNADFYGNQIKPMFCAGYPEGGIDACQDSGGPFCEDSISRTPRKRLCGIVS 376
QY 252 WSGGCAKARPGVGNMVFPTDWYRQMR 280
DB 377 WGTGALAKQKPGVYTKVSDPREMIPOAIK 405

```

```

Search completed: August 1, 2003, 18:13:31
Job time : 21 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:09:49 ; Search time 36 Seconds

(without alignments)
2028.580 Million cell updates/sec

Title: US-09-988-975A-1

Perfect score: 1543

Sequence: 1 MKLNTSAGNVDIYKLYKLYHSD.....YXGNVVFDTWYRQNRADG 283

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP viirus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1540	99.8	492	4	Q96T73
2	1263.5	81.9	490	11	Q920K3
3	764.5	49.5	767	13	Q9NBR2
4	695	45.0	453	11	Q8VDR0
5	694	45.0	453	11	Q8WIT0
6	595	38.6	471	11	Q8CFE0
7	588	38.1	537	4	Q9BEI1
8	587.5	38.1	371	11	Q8CJ16
9	587.5	38.1	445	11	Q8CJ17
10	585	37.9	581	4	Q9BYE2
11	572.5	37.1	455	11	Q8CDR0
12	558.5	36.2	405	4	Q96R86
13	557.5	36.1	435	11	Q8VCA5
14	553	35.8	643	6	Q97506
15	547.5	35.5	624	11	Q9PMT3
16	541.5	35.1	624	11	Q91Y47

17	534	34.6	624	6	Q95ME7	Q95ME7
18	530.5	34.4	638	11	Q8ROP5	Q8ROP5
19	522	33.8	777	11	Q8CAN9	Q8CAN9
20	520	33.7	310	11	Q9QY29	Q9QY29
21	514	33.3	310	11	Q91XC4	Q91XC4
22	512.5	33.2	422	4	Q8WVC1	Q8WVC1
23	511	33.1	799	11	Q9DBI0	Q9DBI0
24	509.5	33.0	855	11	Q93J17	Q93J17
25	507	32.9	435	11	Q8BHM9	Q8BHM9
26	506	32.8	417	11	Q8VHC4	Q8VHC4
27	504.5	32.7	812	11	Q91WJ5	Q91WJ5
28	503	32.6	417	11	Q8VDV1	Q8VDV1
29	502.5	32.6	417	11	Q8VHK8	Q8VHK8
30	502.5	32.6	812	11	Q9ROW3	Q9ROW3
31	500.5	32.4	329	6	Q9GLI0	Q9GLI0
32	500.5	32.4	331	11	Q8RIA6	Q8RIA6
33	500.5	32.4	389	11	Q9PYX7	Q9PYX7
34	496	32.1	328	11	Q8BGR6	Q8BGR6
35	495.5	32.1	257	11	Q8B204	Q8B204
36	495.5	32.1	257	11	Q8B210	Q8B210
37	494.5	32.0	279	11	Q9QZ74	Q9QZ74
38	490	31.8	1524	13	Q91674	Q91674
39	484	31.4	802	4	Q8TUB2	Q8TUB2
40	484	31.4	811	4	Q8TUB0	Q8TUB0
41	481.5	31.2	415	6	Q29015	Q29015
42	480.5	31.1	572	11	Q8B1K6	Q8B1K6
43	478	31.0	845	13	Q9DGR1	Q9DGR1
44	476.5	30.9	416	11	Q8B230	Q8B230
45	476.5	30.9	416	11	Q8B213	Q8B213

ALIGNMENTS

RESULT 1
Q96T73
ID Q96T73 PRELIMINARY; PRT; 492 AA.
AC Q96T73;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
GN Bpitheliasin.
DE TMPSR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223025; PubMed=11322890;
RA Jacquinet E., Rao N.V., Rao G.V., Wang Z., Albertine K.H.,
Hoidal J.R.;
RT "Cloning and characterization of the CDNA and gene for human
epitheliasin".
RT Eur. J. Biochem. 268:2687-2699(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC EMBL; AF329454; AAKS3559.1; -
DR HSSP; P00761; IAN1.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Src_kinase_receptor.
DR Pfam; PF00089; Trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDla; 1.
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS01209; LDla_1; 1.
DR PROSITE; PS00668; LDla_2; 1.
DR PROSITE; PS0287; SRCR_2; 1.
DR PROSITE; PS02440; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.

KW Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 492 AA; 53863 MW; 3ABAT55BP276DAUF CRC64;
 Query Match 99.8%; Score 1540; DB 4; Length 492;
 Best Local Similarity 99.6%; Pred. No. 2,6e-144;
 Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLSSRSQRIVGESALPGAMPW 60
 DB 210 MKLNTAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLSSRSQRIVGESALPGAMPW 269
 QY 61 QVSLHVNQVHVGCGSITTPMTVTAAHCVEKPLNPNMHTAFAGILQSPMPYAGAYQV 120
 DB 270 QVSLHVNQVHVGCGSITTPMTVTAAHCVEKPLNPNMHTAFAGILQSPMPYAGAYQV 329
 QY 121 KVISHPNDYKTKNDIALMKLQKPLTFNDLVKVCCLPNPQMLQPEOLCMIWGKATTE 180
 DB 330 KVISHPNDYKTKNDIALMKLQKPLTFNDLVKVCCLPNPQMLQPEOLCMIWGKATTE 389
 QY 181 KGTSEVLAATAVLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSQGDGGXLVTSK 240
 DB 390 KGTSEVLAATAVLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSQGDGGXLVTSK 449
 QY 241 NNIMWLIQDTSWGSCCAKAYRPGVGNMVFETDWTYQMRADG 283
 DB 450 NNIMWLIQDTSWGSCCAKAYRPGVGNMVFETDWTYQMRADG 492

RESULT 2
 Q920K3 PRELIMINARY; PRT; 490 AA.
 ID Q920K3
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE TMRPSS2.
 GN TMRPSS2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tazuzuki S.;
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB073550; BAB70683.1; -
 DR HSSP; P00761; IANI.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001272; LDL_receptor_A.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR001190; Strc_protease.
 DR Pfam; PF00057; ldl_recept_a; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00192; LDLa; 1.
 DR SMART; SM00020; SR; 1.
 DR SMART; SM00068; LDLRA_2; 1.
 DR PROSITE; PS50287; SRCA_2; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 490 AA; 53518 MW; 2BC691551CAC409A CRC64;

Query Match 81.9%; Score 1263.5; DB 11; Length 490;
 Best Local Similarity 81.6%; Pred. No. 7,6e-117;
 Matches 220; Conservative 22; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLNTAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLSSRSQRIVGESALPGAMPW 60
 DB 209 MKLNTAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLSSRSQRIVGESALPGAMPW 267

QY 61 QVSLHVNQVHVGCGSITTPMTVTAAHCVEKPLNPNMHTAFAGILQSPMPYAGAYQV 120
 DB 268 QVSLHVNQVHVGCGSITTPMTVTAAHCVEKPLNPNMHTAFAGILQSPMPYAGAYQV 327
 QY 121 KVISHPNDYKTKNDIALMKLQKPLTFNDLVKVCCLPNPQMLQPEOLCMIWGKATTE 180
 DB 328 KVISHPNDYKTKNDIALMKLQKPLTFNDLVKVCCLPNPQMLQPEOLCMIWGKATTE 387
 QY 181 KGTSEVLAATAVLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSQGDGGXLVTSK 240
 DB 388 KGTSEVLAATAVLLIETQRCNSRYVDNLITPAMICAGFLQGVNDSQGDGGXLVTSK 447
 QY 241 NNIMWLIQDTSWGSCCAKAYRPGVGNMVFETDWTYQMRADG 282
 DB 448 NNIMWLIQDTSWGSCCAKAYRPGVGNMVFETDWTYQMRADG 489

RESULT 3
 Q9DGR2 PRELIMINARY; PRT; 767 AA.
 ID Q9DGR2
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Embryonic serine protease-2.
 GN XESP-2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20363741; PubMed=10903452;
 RX Yamada K., Takabatake T., Takeshima K.;
 RT "Isolation and characterization of three novel serine protease genes from Xenopus laevis".
 RL Gene 252:209-216 (2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 DR EMBL; AB038497; BAB08217.1; -
 DR HSSP; P00766; ICHG.
 DR MEROPS; S01.049; -
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001272; LDL_receptor_A.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00261; LDLRECEPTOR.
 DR SMART; SM00192; LDLa; 8.
 DR PROSITE; PS01209; LDLA_1; 8.
 DR PROSITE; PS50068; LDLA_2; 2.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 767 AA; 86001 MW; E0566A38796DE96E CRC64;

Query Match 49.5%; Score 764.5; DB 13; Length 767;
 Best Local Similarity 52.3%; Pred. No. 4,9e-67;
 Matches 146; Conservative 35; Mismatches 97; Indels 1; Gaps 1;

QY 2 KLTNTAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLSSRSQRIVGESALPGAMPW 61
 DB 487 KLTNTAGNVDIYKLYHSDACSSKAVVSLRCLACGVNLSSRSQRIVGESALPGAMPW 545
 QY 62 QVSLHVNQVHVGCGSITTPMTVTAAHCVEKPLNPNMHTAFAGILQSPMPYAGAYQV 121
 DB 546 QVSLHVNQVHVGCGSITTPMTVTAAHCVEKPLNPNMHTAFAGILQSPMPYAGAYQV 605
 QY 122 VISHPNYDSTKKNNDIALMKLQKPLTFNDLVKVCCLPNPQMLQPEOLCMIWGKATTE 181

```

Db      606 ITHPEYKSYTDNDIALMKLDEITFGTTTOPVCLPNSGFWMEAGITTMISGSGSYEG 665
Qy      182 GKTSSEVLAARAKVLLLTORCNSRYVDNLITPAMI CAGFLQGVNDSCGDSGGKLVTSKN 241
Db      666 GSVSTLYQAAIPLIDSNVNCOSYVNGQITSMICAGYLGGVDTCGDSGGGLVAKRN 725
Qy      242 NIMWLLGDTSMGSGGAKAYRPGVYGNVFTDMIRQMR 280
Db      726 GTWMLVGDITSMGDCARANKPGVYGNVFTDMIRQMR 764

RESULT 4
08VIDEO
ID 08VIDEO PRELIMINARY; PRT; 453 AA.
AC 08VIDEO;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE TMPRSS3 protein.
GN TMPRSS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Guipponi M., Scamuffa N., Scott H.S., Rosbaker C., Antonarakis S.E.;
RT "Isolation and characterization of the mouse TmpRSS3 gene."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AJ300738; CAC83350.1; -.
DR HSP; P00761; IANI.
DR MGD; MG1:2155445; TmpRSS3.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Src_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; tryptain; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLA_1; 1.
DR PROSITE; PS0068; LDLA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 453 AA; 49529 MW; 2185697DC8781BD3 CRC64;

Query Match 45.0%; Score 695; DB 11; Length 453;
Best Local Similarity 50.0%; Pred. No. 2e-60;
Matches 136; Conservative 38; Mismatches 92; Indels 6; Gaps 4;

```

```

Db      419 SFGIGCAEVNKPQVYTRITSLFDWIHEQLERD 450
Qy      182 GKTSSEVLAARAKVLLLTORCNSRYVDNLITPAMI CAGFLQGVNDSCGDSGGKLVTSKN 241
Db      666 GSVSTLYQAAIPLIDSNVNCOSYVNGQITSMICAGYLGGVDTCGDSGGGLVAKRN 725
Qy      242 NIMWLLGDTSMGSGGAKAYRPGVYGNVFTDMIRQMR 280
Db      726 GTWMLVGDITSMGDCARANKPGVYGNVFTDMIRQMR 764

RESULT 5
08K10
ID 08K10 PRELIMINARY; PRT; 453 AA.
AC 08K10;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Type II transmembrane serine protease.
GN TMPRSS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Guipponi M., Scamuffa N., Scott H.S., Antonarakis S.E.;
RT "Isolation of the mouse TmpRSS3 genomic DNA sequence."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL; AJ429216; CAD22137.1; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Src_receptor.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF00089; tryptain; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR PROSITE; PS01209; LDLA_1; 1.
DR PROSITE; PS0068; LDLA_2; 1.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease; Transmembrane.
SQ SEQUENCE 453 AA; 49491 MW; 1ABCBF10AF631E6 CRC64;

Query Match 45.0%; Score 694; DB 11; Length 453;
Best Local Similarity 49.6%; Pred. No. 2.5e-60;
Matches 135; Conservative 39; Mismatches 92; Indels 6; Gaps 4;

```

```

DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DB Similar to mosaic serine protease (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC042878; AAH42878.1; -.
KW Protease.
FT NON_TER
SQ SEQUENCE 471 AA; 52535 MW; ED58CFE6B7C3BC4 CRC64;

Query Match 38.6%; Score 595; DB 11; Length 471;
Best Local Similarity 43.0%; Pred. No. 1.8e-50;
Matches 120; Conservative 43; Mismatches 102; Indels 14; Gaps 6;

QY 9 NVDIYKLYHSDACSKAVVSLRCIACGVNINSSROSRIVGSEALPGAMPQVSLHVON 68
DB 196 NTLIOESLVSQ-CFSRRVSLQCSHCGLR--AMTGRIVGALVTSSEKMPQVSLHFGT 251
QY 69 VHVGCSITTPEMITVAACV---EKPLNPMWTAAGILRSFMPYAGAYOVERKVIS 124
DB 252 THICGTLIDAGQVLTAAHCFVTRKLEGG--WKYAGTSNHLQLEAA--SIQIIII 306
QY 125 HENVYDKRTQNDIALMKLOKPLTFNDLVKPVCLPNPGMQLQPEOLCWSGKATBERK 183
DB 307 NSNYTDEDDYDIALRLSKPLTSAHHPACLPNHGQTFSLNETCWTIFGKTRTDK 366
QY 184 TSEVLAATAVLLIFORCNSRYVNDLITPAMICAGFLQGVNDSGGDSCGLVTSKNII 243
DB 367 TSPFLREVQVNLIDPKKNDLVYDSYLTTPMCMCAGDLRGDRSCGDSGPIVCEQNNR 426
QY 244 WMLIGDTSGSCCAKAYRPGVGNVFTDWIYRQMBAD 282
DB 427 WYLAGVTSWGTGGGGRNKGVTYKTVLPMYIKMSE 465

RESULT 7
Q9BYE1 PRELIMINARY; PRT; 537 AA.
AC Q9BYE1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Mosaic serine protease.
MSPS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RL MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmir S., Inoue M., Kido H.;
RT "Cloning and expression of novel mosaic serine proteases with and
without a transmembrane domain from human lung.";
BL Biochim. Biophys. Acta 1518:204-209(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL: AB048797; BAB39742.1; -.
DR HSSP: P00763; IDPO.
DR MEROPS: S01.067; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL receptor A.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR001190; Serp_receptor.
DR Pfam: PF00057; Idl_recept_a; 1.

```

```

DR Pfam: PF00089; clypsain; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDla; 1.
DR SMART: SM00022; SR; 1.
DR SMART: SM00020; TYPD_SPC; 1.
DR PROSITE: PSS0287; SERC 2; 1.
DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SRR; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 537 AA; 58102 MW; A39FAE8616DAECF CRC64;

Query Match 38.1%; Score 588; DB 4; Length 537;
Best Local Similarity 42.7%; Pred. No. 1.1e-49;
Matches 119; Conservative 41; Mismatches 105; Indels 14; Gaps 6;

QY 9 NVDIYKLYHSDACSKAVVSLRCIACGVNINSSROSRIVGSEALPGAMPQVSLHVON 68
DB 262 NTLIOESL-HRSHCSQRYISLQCSHCGLR--AMTGRIVGALVTSSEKMPQVSLHFGT 317
QY 69 VHVGCSITTPEMITVAACV---EKPLNPMWTAAGILRSFMPYAGAYOVERKVIS 124
DB 318 THICGTLIDAGQVLTAAHCFVTRKLEGG--WKYAGTSNHLQLEAA--SIQIIII 372
QY 125 HENVYDKRTQNDIALMKLOKPLTFNDLVKPVCLPNPGMQLQPEOLCWSGKATBER 183
DB 373 NSNYTDEDDYDIALRLSKPLTSAHHPACLPNHGQTFSLNETCWTIFGKTRTDK 432
QY 184 TSEVLAATAVLLIFORCNSRYVNDLITPAMICAGFLQGVNDSGGDSCGLVTSKNII 243
DB 433 TSPFLREVQVNLIDPKKNDLVYDSYLTTPMCMCAGDLRGDRSCGDSGPIVCEQNNR 492
QY 244 WMLIGDTSGSCCAKAYRPGVGNVFTDWIYRQMBAD 282
DB 493 WYLAGVTSWGTGGGGRNKGVTYKTVLPMYIKMSE 531

RESULT 8
Q8CJ16 PRELIMINARY; PRT; 371 AA.
AC Q8CJ16;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Adrenal mitochondrial protease short variant.
GN AHP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NEBH;
RA Omer S., Bicknell A.B., Lowry P.U.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF537099; AAN06758.1; -.
KW Protease.
SQ SEQUENCE 371 AA; 40694 MW; 89A64081D9A1FE26 CRC64;

Query Match 38.1%; Score 587.5; DB 11; Length 371;
Best Local Similarity 42.7%; Pred. No. 7.4e-50;
Matches 111; Conservative 39; Mismatches 105; Indels 5; Gaps 3;

QY 19 SDACSKAVVSLRCIACGVNINSSROSRIVGSEALPGAMPQVSLHVONVHVGCSIT 78
DB 109 STNCPGSRIVSLKCEGQAR--PLASRIVGQVAGVGRMPQVSLHSGSRHICGGSVLA 165
QY 79 PEMITVAACVCKR-FLNPMWTAAGILRSFMPYAGAYOVERKVISHPYDRTKQNDI 137
DB 166 PIVVVAHACQTSFRLSRSSRRVYAGVLSAVARHOGTVEKCLIPHPYLSKQNDHYDV 225
QY 138 ALMKLOKPLTFNDLVKPVCLPNPGMQLQPEOLCWSGKATBERKGTSEVLAATAVLLI 196

```

```

DB 226 ALAQTRPSTVAVCLPAKEQHFPQSGQCMVSGMHTDPSHTSSDTLQDTWPL 285
QY 197 ETQRCNSRYVDNLITPAMICAGFLQGNVDSQCGSGXLYTSKNNIMWLIGDTSWGGC 256
DB 286 STDLCSSCMYSGALTHRLCAGYLDGRADACQCGSGGLVCPSGDTWHLVGVSWGRGC 345
QY 257 AKAYRPGVYGNVWFTDWY 276
DB 346 AEPNRPGVYAKVAEFLDWIH 365

RESULT 9
Q8CJ17 PRELIMINARY; PRT; 445 AA.
AC Q8CJ17;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Adrenal mitochondrial protease long variant.
GN AMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEBH;
RA Omer S., Bicknell A.B., Lowy P.J.;
RT "Identification of a rat adrenal mitochondrial protease.";
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
KW EMBL: A5537098; ANM06757.1; -.
SQ SEQUENCE 445 AA; 48440 MW; BE3F56D8372ED988 CRC64;

Query Match 38.1%; Score 587.5; DB 11; Length 445;
Best Local Similarity 42.7%; Pred. No. 9.3e-50;
Matches 111; Conservative 39; Mismatches 105; Indels 5; Gaps 3;

QY 19 SDACSKAVVSLRCLACGYNLNLSRQSRIVGGESALPGAMFWQVSLHVNVAHVGGSIT 78
DB 183 STNCPGRIVSLKCECGAR---PLASRIVGGAVASGRWQASVMLGSRHTCGGSVLA 239
QY 79 PENIVTAACVEK-PLANPMHTAFAGILROSFMYGAGYOVEKVISHPNYSKTKNDI 137
DB 240 PYWVVAACWYSFRLSRLSRWVHAGLVSHSAVQHOGTWVEKLIHPILYSQNDHYD 239
QY 138 ALMKLQKPLTNDLVKPVCLPMPGMLOEBOLCWISGKATE-EKAKTSEVLAATVLLI 196
DB 300 ALIQETPTNFSDTVASVCLPAKEQHFPQSGQCMVSGMHTDPSHTSSDTLQDTWPL 359
QY 197 ETQRCNSRYVDNLITPAMICAGFLQGNVDSQCGSGXLYTSKNNIMWLIGDTSWGGC 256
DB 360 STDLCSSCMYSGALTHRLCAGYLDGRADACQCGSGGLVCPSGDTWHLVGVSWGRGC 419
QY 257 AKAYRPGVYGNVWFTDWY 276
DB 420 AEPNRPGVYAKVAEFLDWIH 439

RESULT 10
Q8BYE2 PRELIMINARY; PRT; 581 AA.
AC Q8BYE2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Membrane-type 23, Last annotation update)
GN MSPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=21167393; PubMed=11267681;
RA Kim D.R., Sharmila S., Inoue M., Kido H.;
RT "Cloning and expression of novel mouse serine proteases with and
RT without a transmembrane domain from human lung.";
RL Biochim. Biophys. Acta 1518:204-209(2001).
DB EMBL: AB048796; BAB39741.1; -.
DR HSP; P00763; IDPO.
DR MEROPS; S01.087; -.
DR InterPro; IPR002172; IDL_receptor_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Srcf_receptor.
DR Pfam; PF00057; Idl_recept_a; 1.
DR Pfam; PF00089; Trypsin; 1.
DR SMART; SM00192; LDla; 1.
DR SMART; SM00202; SR; 1.
DR SMART; SM00202; TRYP_SPC; 1.
DR PROSITE; PS02027; SRCR_2; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Protease; Serine protease.
SQ SEQUENCE 581 AA; 62689 MW; 4DABE24D7D5BA44 CRC64;

Query Match 37.9%; Score 585; DB 4; Length 581;
Best Local Similarity 43.1%; Pred. No. 2.3e-49;
Matches 119; Conservative 39; Mismatches 104; Indels 14; Gaps 6;

QY 9 NVDIYKLVHSNACSKAVVSLRCLACGYNLNLSRQSRIVGGESALPGAMFWQVSLHVN 68
DB 287 NSTIQEBSL-HRSHCPQSRVYSLQCSHCGLR---AMTGRVGGALASDSKMPWQVSLHFGT 342
QY 69 VHVCGSIIITPEWITVAACV---EKPLANPMHTAFAGILROSFMYGAGYOVEKYS 124
DB 343 THICGLIDLAQVNLPAKCFPTREKYLEG---KRYVAGTSNHLQLEAA-SIAEIT 397
QY 125 HPNYDSKTKNDIATMLKQKPLTNDLVKPVCLPMPGMLOEBOLCWISGKATEE-KGK 183
DB 398 NSNTDEEDDYDALMRLSKPLTSLAHIPACLPMHGQTFSLNRCWITGCKREITDK 457
QY 184 TSEVLAATVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSQCGSGXLYTSKNNI 243
DB 458 TSPFLARVQVNLIDPKKNDLYVDSYLTFRMCKAGDLHGRDSQCGSGGLVCEQNNR 517
QY 244 WMLIGDTSWGGCAKAYRPGVYGNVWFTDWY 279
DB 518 WYLAGVTSWGTGCGQRRKQGVYKTEVLPWYISQK 553

RESULT 11
Q8CDRO PRELIMINARY; PRT; 455 AA.
AC Q8CDRO;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Transmembrane protease.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RT the Riken Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL: AK029714; BAC26577.1; -.

```

SQ SEQUENCE 455 AA; 49669 MW; B822EB2E7503C74B CRC64;
Query Match 37.1%; Score 572.5; DB 11; Length 455;
Best Local Similarity 42.0%; Pred. No. 3e-48;
Matches 108; Conservative 39; Mismatches 105; Indels 5; Gaps 3;
QY 22 CASKAVSLRCLACGVNINSSRQSHVGSALNGANPWQVSLVQNVHVCSSITTPW 81
DB 196 CFSGRIVSLKCECCAR--PLASRIVGGAQVAGRWPMQASVNLGSRHTCGASVLAIPH 252
QY 82 IYTAHCVCK-PLANNPWHTAFAGILRQSPFMYGAGYQVEKVISHPNYSKTKNDIALM 140
DB 253 VYTAHCHVSPFLSLSLSSMRVHAGVSHGAVRQHGTMVEKLLIPHLVSAQNDYDVAL 312
QY 141 KQKPLTFNDLYKPYCLPRPGMMLQPEQLCWTISGKATE-EKKTSTEVLAAYVLLIEQ 199
DB 313 QLRTPNNSDVTGAVCALCPKQGHFPMWSQCVWSGWHGTHSHSDTLODTWVPLLSY 372
QY 200 RGNRHYVDNLITTPAMICAGFLQGVNDSQGDSCGLVTSKNNIMWLIIGDTSWGSCAKA 259
DB 373 LCNSSGCMISGALTHMLCAGYLDGRADACQDSGGPLVCP89GDTWHLVGVVSGRCAP 432
QY 260 YRPGYGVNMTWTDWY 276
DB 433 NRPVYAKVAEFLDMIH 449
RESULT 12
Q96E86 PRELIMINARY; PRT; 405 AA.
ID Q96E86;
AC 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strauberg R.;
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL: BC012752; AAI12752.1; --
DR HSP: P00761; IANL.
DR MEROPS: S01.034; --
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001212; IDL_receptor_A.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR001190; Strc_receptor.
DR Pfam: PF00057; Idl_recept_a; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR SMART: SM00192; IDLa; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00202; TRYP_SPC; 1.
DR PROSITE: PS00287; SRCR 2; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease; Transmembrane.
FT NON_TER
SQ SEQUENCE 405 AA; 44474 MW; 951ACD52D9D48E04 CRC64;
Query Match 36.2%; Score 558.5; DB 4; Length 405;
Best Local Similarity 43.9%; Pred. No. 6.3e-47;
Matches 118; Conservative 41; Mismatches 93; Indels 17; Gaps 8;
QY 19 SDACSKAVVSLRCLACGVNINSSRQSHVGSALNGANPWQVSLVQNVHVCSSIT 78
DB 148 SGPCLSGSLVSLKCLDCGKSL--KTPRVAGVEAPVDSWPMQVSIQYRKHVCAGSILD 204

QY 79 PEVITVAHCVCKEPLNNPWHTAFAGILR-QSPFMYGAGYQVEKVI---SHPNYSKTKN 134
DB 205 PHVNLRAACPFRRH-TDVNWKVRAGSDKLSP-----PSLAVAKIILIERNPMV---PKD 256
QY 135 NDIALMKLQKPLTFNDLYKPYCLPRPGMMLQPEQLCWTISGKATEEK-GKTSVLAAY 193
DB 257 NDIALMKLQKPLTFNDLYKPYCLPRPGMMLQPEQLCWTISGKATEEK-GKTSVLAAY 316
QY 194 LLIETORCNSRYVDNLITTPAMICAGFLQGVNDSQGDSCGLVTSKNNIMWLIIGDTSWG 253
DB 317 QVIDSTRCNADADYQGEVTEKMKACGIRPEGVDVCCQDSGGPLVYQSDQ--HVVYGIYSWG 375
QY 254 SGCARVYGVNMTWTDWY 282
DB 376 YGCGPSTPGVYTKVASVLAHMTYVWKAE 404

RESULT 13
Q8VCAS PRELIMINARY; PRT; 435 AA.
ID Q8VCAS;
AC 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Similar to transmembrane protease, serine 4 (Channel-activating protease 2).
GN TMPS84.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strauberg R.;
RL Submitted (Jan-2002) to the EMBL/Genbank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=2214321; PubMed=12149280;
RA Veuclin G., Vallet V., Jaeger N.F., Hummler E., Roessler B.C.;
RT "Synergistic Activation of ENaC by Three Membrane-bound Channel-
activating Serine Proteases (mCAP1, mCAP2, and mCAP3) and Serum- and
RT Glucocorticoid-regulated Kinase (SGK1) in Xenopus Oocytes.",
RT J. Gen. Physiol. 120:191-201(2002).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR EMBL: BC021368; AAH21368.1; --
DR EMBL: AY043240; AAK85307.1; --
DR HSP: P00761; IANL.
DR MEROPS: S01.034; --
DR MED: MGI:2384877; TMPS84.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001212; IDL_receptor_A.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR001190; Strc_receptor.
DR Pfam: PF00057; Idl_recept_a; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR PROSITE: PS00287; SRCR 2; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease; Transmembrane.
FT NON_TER
SQ SEQUENCE 435 AA; 47495 MW; DC52BA5A43E01369 CRC64;
Query Match 36.1%; Score 557.5; DB 11; Length 435;
Best Local Similarity 43.8%; Pred. No. 8.7e-47;
Matches 117; Conservative 41; Mismatches 96; Indels 13; Gaps 7;
QY 19 SDACSKAVVSLRCLACGVNINSSRQSHVGSALNGANPWQVSLVQNVHVCSSIT 78
DB 178 SRSCLSGSLVSLKCLDCGKSL--KTPRVAGVEAPVDSWPMQVSIQYRKHVCAGSILD 234

DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 624 AA; 69788 MW; 0ERDDBC56009E97 CRC64;

Query Match 35.5%; Score 547.5; DB 11; Length 624;
 Best Local Similarity 42.1%; Pred. No. 1.3e-45;
 Matches 102; Conservative 43; Mismatches 90; Indels 7; Gaps 4;

QY	35	CGVNINSSRQSRIVGGESALPGAMPQVSLHVQNVHVCQGSITPEWIVTAHCVKPLN	94
Db	382	CTTKIN---PRVGGASVHGEMPMOVLHISQGLCGSIIIGNQWILTAHCF-SGIE	436
QY	95	NPWHWTAFAGILPQSFMYG-AGYQVEKVISHPNDSTKNDIALMKLOKPLTENDLVK	153
Db	437	TPKGLRVYGGIVNQSEINBGTAFRRVQEMIHDOYTTAESGYDIALKLKESAMVTD	496
QY	154	PVCLPMPGMLOPBOELCWSGWATEBKKTSEVLNAKVLLEFORCNSRYVDNLITP	213
Db	497	PICLPKSGDRNAVHTECWTGMYTALRGVOSTLOKAKVPLVSNESCQTRY-RRHKITN	555
QY	214	AMTCAGPLQANDSCQSGSGSLVTSKXNIMWLIGDTSMGSCAKAYRPGVYGVWVFTD	273
Db	556	KMICAGIKREGKQDKGDSGGLPSCCKNGWMLVGIISWEGCGKERPGVITVAKYVD	615
QY	274	WI 275	
Db	616	WI 617	

Search completed: August 1, 2003, 18:12:41
 Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: August 1, 2003, 18:12:44 ; Search time 52 Seconds
(without alignments)

646.327 Million cell updates/sec

Title: US-09-988-975A-1

Perfect score: 1543

Sequence: 1 MKNTSAGNVDIYKLYHSD.....VYGNVWFPTWYRQMRADG 283

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09A_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1541	99.9	283 10 US-09-988-975A-1	Sequence 1, App1
2	1541	99.9	283 15 US-10-235-699-1	Sequence 1, App1
3	1540	99.8	393 9 US-09-759-143-934	Sequence 934, App
4	1540	99.8	393 9 US-09-780-669-934	Sequence 934, App
5	1540	99.8	393 9 US-09-822-827-934	Sequence 934, App
6	1540	99.8	393 10 US-09-895-793-934	Sequence 934, App
7	1540	99.8	393 10 US-09-895-814-934	Sequence 934, App
8	1540	99.8	393 14 US-10-012-896-934	Sequence 934, App
9	1540	99.8	492 9 US-09-759-143-932	Sequence 932, App
10	1540	99.8	492 9 US-09-780-669-932	Sequence 932, App
11	1540	99.8	492 9 US-09-822-827-932	Sequence 932, App
12	1540	99.8	492 10 US-09-895-793-932	Sequence 932, App
13	1540	99.8	492 10 US-09-895-814-932	Sequence 932, App
14	1540	99.8	492 10 US-10-012-896-932	Sequence 932, App
15	1540	99.8	492 15 US-10-205-823-415	Sequence 415, App

16	1535	99.5	384 10 US-09-981-353-23	Sequence 23, App1
17	1522	98.6	492 9 US-09-759-143-895	Sequence 895, App
18	1522	98.6	492 9 US-09-780-669-895	Sequence 895, App
19	1522	98.6	492 9 US-09-879-792-14	Sequence 14, App1
20	1522	98.6	492 9 US-09-822-827-895	Sequence 895, App
21	1522	98.6	492 10 US-09-895-793-895	Sequence 895, App
22	1522	98.6	492 10 US-09-895-814-895	Sequence 895, App
23	1522	98.6	492 11 US-09-776-191-70	Sequence 70, App1
24	1522	98.6	492 14 US-10-012-896-895	Sequence 895, App
25	1254	81.3	229 11 US-09-978-837A-53	Sequence 53, App1
26	778.5	50.5	418 15 US-10-177-661-6	Sequence 6, App1
27	696	45.1	453 10 US-09-978-295A-69	Sequence 69, App1
28	696	45.1	453 10 US-09-978-697-69	Sequence 69, App1
29	696	45.1	453 10 US-09-978-192A-69	Sequence 69, App1
30	696	45.1	453 10 US-09-999-832A-69	Sequence 69, App1
31	696	45.1	453 11 US-09-978-189-69	Sequence 69, App1
32	696	45.1	453 11 US-09-978-608A-69	Sequence 69, App1
33	696	45.1	453 11 US-09-978-585A-69	Sequence 69, App1
34	696	45.1	453 11 US-09-978-191A-69	Sequence 69, App1
35	696	45.1	453 11 US-09-978-403A-69	Sequence 69, App1
36	696	45.1	453 11 US-09-978-564A-69	Sequence 69, App1
37	696	45.1	453 11 US-09-999-833A-69	Sequence 69, App1
38	696	45.1	453 11 US-09-981-915A-69	Sequence 69, App1
39	696	45.1	453 11 US-09-978-824-69	Sequence 69, App1
40	696	45.1	453 11 US-09-918-585A-69	Sequence 69, App1
41	696	45.1	453 11 US-09-978-423A-69	Sequence 69, App1
42	696	45.1	453 11 US-09-978-193A-69	Sequence 69, App1
43	696	45.1	453 11 US-09-999-830A-69	Sequence 69, App1
44	696	45.1	453 11 US-09-978-757A-69	Sequence 69, App1
45	696	45.1	453 11 US-09-978-187B-69	Sequence 69, App1

ALIGNMENTS

RESULT 1
US-09-988-975A-1
Sequence 1, Application US/09988975A
Patent No. US20020119531A1
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Iai, Preeti G.
TITLE OF INVENTION: PROSTATE-ASSOCIATED PROTEIN ANTIBODY
FILE REFERENCE: PR-0227-2 CIP
CURRENT APPLICATION NUMBER: US/09/988, 975A
CURRENT FILING DATE: 2001-11-19
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 283
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020119531A1 556016
FEATURE:
NAME/KEY: unsure
LOCATION: 235
OTHER INFORMATION: unknown or other
US-09-988-975A-1

Query Match 99.9%; Score 1541; DB 10; Length 283;
Best Local Similarity 100.0%; Pred. No. 2.5e-147;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNTSAGNVDIYKLYHSDACSSRAVSLKICAGVINSRSRIVGSSALPGAMPW 60
DB 1 MKNTSAGNVDIYKLYHSDACSSRAVSLKICAGVINSRSRIVGSSALPGAMPW 60
QY 61 QVSLHVNVAVCGSIIITPEWITVAHCEKPLNPMWMTAFAGILROSFFETGAGYQVE 120
DB 61 QVSLHVNVAVCGSIIITPEWITVAHCEKPLNPMWMTAFAGILROSFFETGAGYQVE 120

QY 121 KVISHPYDSKTKNDIAMLKLOKPLTFNDLVKPVCLNPGMLOPEOLCWSGMAATEE 180
DB 121 KVISHPYDSKTKNDIAMLKLOKPLTFNDLVKPVCLNPGMLOPEOLCWSGMAATEE 180
QY 181 KGTSEVLAANAAYLLIETORCNSRYVYDNLITPAMICAGFLOGVNDSCQDGGXLYTSK 240
DB 181 KGTSEVLAANAAYLLIETORCNSRYVYDNLITPAMICAGFLOGVNDSCQDGGXLYTSK 240
QY 241 NNIMWLLIGDTSWGSOGCAKAYRPGVGNVWFTDWTYRQMRADG 283
DB 241 NNIMWLLIGDTSWGSOGCAKAYRPGVGNVWFTDWTYRQMRADG 283

RESULT 2
US-10-235-699-1

/ Sequence 1, Application US/10235699
/ Publication No. US20030103981A1
/ GENERAL INFORMATION:
/ APPLICANT: Spencake, Kimberly M.
/ APPLICANT: Bandman, Olga
/ APPLICANT: Lal, Preeti G.
/ TITLE OF INVENTION: METHOD OF USE OF A PROSTATE-ASSOCIATED PROTEASE IN THE DIAGNOSIS
/ TITLE OF INVENTION: TREATMENT OF PROSTATE CANCER
/ FILE REFERENCE: PV-0010 CIP
/ CURRENT APPLICATION NUMBER: US/10/235,699
/ CURRENT FILING DATE: 2002-09-04
/ PRIOR APPLICATION NUMBER: 09/988,975
/ PRIOR FILING DATE: 11/19/01
/ PRIOR APPLICATION NUMBER: 09/478,957
/ PRIOR FILING DATE: 01/07/00
/ PRIOR APPLICATION NUMBER: 08/807,151
/ PRIOR FILING DATE: 02/27/97
/ NUMBER OF SEQ ID NOS: 8
/ SOFTWARE: PERL Program
/ SEQ ID NO 1
/ LENGTH: 283
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc.feature
/ OTHER INFORMATION: Incyte ID No. US20030103981A1 556016CD1
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 235
/ OTHER INFORMATION: unknown or other
US-10-235-699-1

Query Match 99.9%; Score 1541; DB 15; Length 283;
Best Local Similarity 100.0%; Pred. No. 2.5e-147;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNTSAGNDIYKIKYHSDACSSKAVVSLRCLACGVNLSRSGRIYVGSALPGAMP 60
DB 1 MKNTSAGNDIYKIKYHSDACSSKAVVSLRCLACGVNLSRSGRIYVGSALPGAMP 60
QY 61 QVSLHVNVAHVCSSITTPBWTYTAHCVCKPLNPMWMTAFAGILRSFMYGAGYVE 120
DB 61 QVSLHVNVAHVCSSITTPBWTYTAHCVCKPLNPMWMTAFAGILRSFMYGAGYVE 120
QY 121 KVISHPYDSKTKNDIAMLKLOKPLTFNDLVKPVCLNPGMLOPEOLCWSGMAATEE 180
DB 121 KVISHPYDSKTKNDIAMLKLOKPLTFNDLVKPVCLNPGMLOPEOLCWSGMAATEE 180
QY 181 KGTSEVLAANAAYLLIETORCNSRYVYDNLITPAMICAGFLOGVNDSCQDGGXLYTSK 240
DB 181 KGTSEVLAANAAYLLIETORCNSRYVYDNLITPAMICAGFLOGVNDSCQDGGXLYTSK 240
QY 241 NNIMWLLIGDTSWGSOGCAKAYRPGVGNVWFTDWTYRQMRADG 283
DB 241 NNIMWLLIGDTSWGSOGCAKAYRPGVGNVWFTDWTYRQMRADG 283

RESULT 3

US-09-759-143-934
/ Sequence 934, Application US/09759143
/ Patent No. US20020022248A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Dillon, Devin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vecivick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel
/ APPLICANT: Wang, Aljun
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Hepler, William
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
/ FILE REFERENCE: 210121.427C23
/ CURRENT APPLICATION NUMBER: US/09/759,143
/ CURRENT FILING DATE: 2001-01-12
/ NUMBER OF SEQ ID NOS: 934
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 934
/ LENGTH: 393
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-759-143-934

Query Match 99.8%; Score 1540; DB 9; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKNTSAGNDIYKIKYHSDACSSKAVVSLRCLACGVNLSRSGRIYVGSALPGAMP 60
DB 111 MKNTSAGNDIYKIKYHSDACSSKAVVSLRCLACGVNLSRSGRIYVGSALPGAMP 170
QY 61 QVSLHVNVAHVCSSITTPBWTYTAHCVCKPLNPMWMTAFAGILRSFMYGAGYVE 120
DB 171 QVSLHVNVAHVCSSITTPBWTYTAHCVCKPLNPMWMTAFAGILRSFMYGAGYVE 230
QY 121 KVISHPYDSKTKNDIAMLKLOKPLTFNDLVKPVCLNPGMLOPEOLCWSGMAATEE 180
DB 231 KVISHPYDSKTKNDIAMLKLOKPLTFNDLVKPVCLNPGMLOPEOLCWSGMAATEE 290
QY 181 KGTSEVLAANAAYLLIETORCNSRYVYDNLITPAMICAGFLOGVNDSCQDGGXLYTSK 240
DB 291 KGTSEVLAANAAYLLIETORCNSRYVYDNLITPAMICAGFLOGVNDSCQDGGXLYTSK 350
QY 241 NNIMWLLIGDTSWGSOGCAKAYRPGVGNVWFTDWTYRQMRADG 283
DB 351 NNIMWLLIGDTSWGSOGCAKAYRPGVGNVWFTDWTYRQMRADG 393

RESULT 4
US-09-780-669-934

/ Sequence 934, Application US/09780669
/ Patent No. US2002005197A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Dillon, Devin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.

```

APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedrick, Thomas S.
APPLICANT: Carter, Derrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.427C24
CURRENT APPLICATION NUMBER: US/09/780,669
NUMBER OF SEQ ID NOS: 943
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-780-669-934

```

```

Query Match          99.8%; Score 1540; DB 9; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 MKLNTSAGNDIYKYLHSDACSSKAVVSLRCLACGVLNNSRQRIYVGSALPGAMPW 60
Db 111 MKLNTSAGNDIYKYLHSDACSSKAVVSLRCLACGVLNNSRQRIYVGSALPGAMPW 170
Qy 61 QVSLHVNQVHVCSSITTPWITVAHCVKELNPMWHTAFAGILROSFMFYGAGYQVE 120
Db 171 QVSLHVNQVHVCSSITTPWITVAHCVKELNPMWHTAFAGILROSFMFYGAGYQVE 230
Qy 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLQPEOLCWSGKATEE 180
Db 231 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLQPEOLCWSGKATEE 290
Qy 181 KGTSEVLNAKVLLETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Db 291 KGTSEVLNAKVLLETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 350
Qy 241 NNTMMLIGDTSMGSCAKAYRPGYGVNMFDTMIYRORADG 283
Db 351 NNTMMLIGDTSMGSCAKAYRPGYGVNMFDTMIYRORADG 393

```

```

RESULT 5
US-09-822-827-934
Sequence 934, Application US/09822827
Patent No. US20020081680A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C1
CURRENT APPLICATION NUMBER: US/09/822,827
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-822-827-934

```

```

Query Match          99.8%; Score 1540; DB 9; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MKLNTSAGNDIYKYLHSDACSSKAVVSLRCLACGVLNNSRQRIYVGSALPGAMPW 60

```

```

Db 111 MKLNTSAGNDIYKYLHSDACSSKAVVSLRCLACGVLNNSRQRIYVGSALPGAMPW 170
Qy 61 QVSLHVNQVHVCSSITTPWITVAHCVKELNPMWHTAFAGILROSFMFYGAGYQVE 120
Db 171 QVSLHVNQVHVCSSITTPWITVAHCVKELNPMWHTAFAGILROSFMFYGAGYQVE 230
Qy 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLQPEOLCWSGKATEE 180
Db 231 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLQPEOLCWSGKATEE 290
Qy 181 KGTSEVLNAKVLLETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 240
Db 291 KGTSEVLNAKVLLETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGLVTSK 350
Qy 241 NNTMMLIGDTSMGSCAKAYRPGYGVNMFDTMIYRORADG 283
Db 351 NNTMMLIGDTSMGSCAKAYRPGYGVNMFDTMIYRORADG 393

```

```

RESULT 6
US-09-895-793-934
Sequence 934, Application US/09895793
Publication No. US20020192763A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kaios, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedrick, Thomas S.
APPLICANT: Carter, Derrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Baeseols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.534C2
CURRENT APPLICATION NUMBER: US/09/895,793
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 934
LENGTH: 393
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-793-934

```

```

Query Match          99.8%; Score 1540; DB 10; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 MKLNTSAGNDIYKYLHSDACSSKAVVSLRCLACGVLNNSRQRIYVGSALPGAMPW 60
Db 111 MKLNTSAGNDIYKYLHSDACSSKAVVSLRCLACGVLNNSRQRIYVGSALPGAMPW 170
Qy 61 QVSLHVNQVHVCSSITTPWITVAHCVKELNPMWHTAFAGILROSFMFYGAGYQVE 120
Db 171 QVSLHVNQVHVCSSITTPWITVAHCVKELNPMWHTAFAGILROSFMFYGAGYQVE 230
Qy 121 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCLPFGMMLQPEOLCWSGKATEE 180

```

```
Db 231 |||||
Qy 181 KGTSEVLAAGVLLIETORCNSRYVYDNLITPAMICAGFLQGVDSGQDSGXLYTSK 240
Db 291 KGTSEVLAAGVLLIETORCNSRYVYDNLITPAMICAGFLQGVDSGQDSGXLYTSK 350
Qy 241 NNIMWLIGDTSWGSACAKAYRPGVYGNVVFPTDWTYROMRADG 283
Db 351 NNIMWLIGDTSWGSACAKAYRPGVYGNVVFPTDWTYROMRADG 393

RESULT 7
US-09-895-814-934
; Sequence 934, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugui
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassole, Carlota
; APPLICANT: Roy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ. ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PR1
; ORGANISM: Homo sapiens
; US-09-895-814-934

Query Match 99.8%; Score 1540; DB 10; Length 393;
Best Local Similarity 99.8%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCIAGVNLSSRSRIYVGSALPGAMP 60
Db 111 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCIAGVNLSSRSRIYVGSALPGAMP 170
Qy 61 QVSLHYQNVHVCSSITTEBMTVTAHCEKPLNPMWMTAFAGILQSFMYGAGYVE 120
Db 171 QVSLHYQNVHVCSSITTEBMTVTAHCEKPLNPMWMTAFAGILQSFMYGAGYVE 230
Qy 121 KVTSHPNYDSTKNDIALMKLQKPLTFNDLVKPYCLNPGMLOPBLQWISGMATTE 180
Db 231 KVTSHPNYDSTKNDIALMKLQKPLTFNDLVKPYCLNPGMLOPBLQWISGMATTE 290
Qy 181 KGTSEVLAAGVLLIETORCNSRYVYDNLITPAMICAGFLQGVDSGQDSGXLYTSK 240
Db 291 KGTSEVLAAGVLLIETORCNSRYVYDNLITPAMICAGFLQGVDSGQDSGXLYTSK 350
Qy 241 NNIMWLIGDTSWGSACAKAYRPGVYGNVVFPTDWTYROMRADG 283
```

```
Db 351 NNIMWLIGDTSWGSACAKAYRPGVYGNVVFPTDWTYROMRADG 393

RESULT 8
US-10-012-896-934
; Sequence 934, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugui
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassole, Carlota
; APPLICANT: Roy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ. ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 934
; LENGTH: 393
; TYPE: PR1
; ORGANISM: Homo sapiens
; US-10-012-896-934

Query Match 99.8%; Score 1540; DB 14; Length 393;
Best Local Similarity 99.6%; Pred. No. 4.8e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCIAGVNLSSRSRIYVGSALPGAMP 60
Db 111 MKLNTSAGNVDIYKCLYHSDACSSKAVVSLRCIAGVNLSSRSRIYVGSALPGAMP 170
Qy 61 QVSLHYQNVHVCSSITTEBMTVTAHCEKPLNPMWMTAFAGILQSFMYGAGYVE 120
Db 171 QVSLHYQNVHVCSSITTEBMTVTAHCEKPLNPMWMTAFAGILQSFMYGAGYVE 230
Qy 121 KVTSHPNYDSTKNDIALMKLQKPLTFNDLVKPYCLNPGMLOPBLQWISGMATTE 180
Db 231 KVTSHPNYDSTKNDIALMKLQKPLTFNDLVKPYCLNPGMLOPBLQWISGMATTE 290
Qy 181 KGTSEVLAAGVLLIETORCNSRYVYDNLITPAMICAGFLQGVDSGQDSGXLYTSK 240
Db 291 KGTSEVLAAGVLLIETORCNSRYVYDNLITPAMICAGFLQGVDSGQDSGXLYTSK 350
Qy 241 NNIMWLIGDTSWGSACAKAYRPGVYGNVVFPTDWTYROMRADG 283
Db 351 NNIMWLIGDTSWGSACAKAYRPGVYGNVVFPTDWTYROMRADG 393

RESULT 9
US-09-759-143-932
```

```
/ Sequence 932, Application US/09759143
/ Patent No. US2002002248A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Dillon, Devin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skelky, Yasir A.W.
/ APPLICANT: Hepler, William
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.427C23
/ CURRENT APPLICATION NUMBER: US/09/759,143
/ CURRENT FILING DATE: 2001-01-12
/ NUMBER OF SEQ ID NOS: 934
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 932
/ LENGTH: 492
/ TYPE: PR
/ ORGANISM: Homo sapiens
US-09-759-143-932

Query Match 99.8%; Score 1540; DB 9; Length 492;
Best Local Similarity 99.6%; Pred. No. 6.4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

/ 1 MKLNTSAGNVDIYKLYHSDACSSKAVYSLRCLACGVNLNSRSGRIYGGSSALPGAMPW 60
/ 210 MKLNTSAGNVDIYKLYHSDACSSKAVYSLRCLACGVNLNSRSGRIYGGSSALPGAMPW 269
/ 61 QVSLHVNQVAVCGGSIITPEWIVTAHCEVEKPLNPMHTAFAGILRQSFNFGAGYQVE 120
/ 270 QVSLHVNQVAVCGGSIITPEWIVTAHCEVEKPLNPMHTAFAGILRQSFNFGAGYQVE 329
/ 121 KYLSHPYDSKTKNDIALMKLOKPLTFNDLVKPYCLNPGMLOPEOLCWSMGATBE 180
/ 330 KYLSHPYDSKTKNDIALMKLOKPLTFNDLVKPYCLNPGMLOPEOLCWSMGATBE 389
/ 181 KKTSEVLNAKYLILETORCNSRYVDNLITPAMICAGFLQGVNDSGQDGGGLVTSK 240
/ 390 KKTSEVLNAKYLILETORCNSRYVDNLITPAMICAGFLQGVNDSGQDGGGLVTSK 449
/ 241 NNIMWLIQDTSWGGCAKAYRPGVYGNVWFTDMLYRMRADG 283
/ 450 NNIMWLIQDTSWGGCAKAYRPGVYGNVWFTDMLYRMRADG 492

RESULT 10
US-09-780-669-932
/ Sequence 932, Application US/09780669
/ Patent No. US2002005197A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Dillon, Devin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
```

```
/ APPLICANT: Day, Craig H.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Carter, Darrick
/ APPLICANT: Li, Samuel
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skelky, Yasir A.W.
/ APPLICANT: Hepler, William
/ APPLICANT: Hurai, John
/ APPLICANT: McNeill, Patricia D.
/ APPLICANT: Houghton, Raymond L.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.427C24
/ CURRENT APPLICATION NUMBER: US/09/780,669
/ CURRENT FILING DATE: 2001-02-09
/ NUMBER OF SEQ ID NOS: 943
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 932
/ LENGTH: 492
/ TYPE: PR
/ ORGANISM: Homo sapiens
US-09-780-669-932
```

```
Query Match 99.8%; Score 1540; DB 9; Length 492;
Best Local Similarity 99.6%; Pred. No. 6.4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
/ 1 MKLNTSAGNVDIYKLYHSDACSSKAVYSLRCLACGVNLNSRSGRIYGGSSALPGAMPW 60
/ 210 MKLNTSAGNVDIYKLYHSDACSSKAVYSLRCLACGVNLNSRSGRIYGGSSALPGAMPW 269
/ 61 QVSLHVNQVAVCGGSIITPEWIVTAHCEVEKPLNPMHTAFAGILRQSFNFGAGYQVE 120
/ 270 QVSLHVNQVAVCGGSIITPEWIVTAHCEVEKPLNPMHTAFAGILRQSFNFGAGYQVE 329
/ 121 KYLSHPYDSKTKNDIALMKLOKPLTFNDLVKPYCLNPGMLOPEOLCWSMGATBE 180
/ 330 KYLSHPYDSKTKNDIALMKLOKPLTFNDLVKPYCLNPGMLOPEOLCWSMGATBE 389
/ 181 KKTSEVLNAKYLILETORCNSRYVDNLITPAMICAGFLQGVNDSGQDGGGLVTSK 240
/ 390 KKTSEVLNAKYLILETORCNSRYVDNLITPAMICAGFLQGVNDSGQDGGGLVTSK 449
/ 241 NNIMWLIQDTSWGGCAKAYRPGVYGNVWFTDMLYRMRADG 283
/ 450 NNIMWLIQDTSWGGCAKAYRPGVYGNVWFTDMLYRMRADG 492

RESULT 11
US-09-822-827-932
/ Sequence 932, Application US/09822827
/ Patent No. US20020081680A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu, Jiangchun
/ APPLICANT: Dillon, Devin C.
/ APPLICANT: Mitcham, Jennifer L.
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Jiang, Yugu
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Stolk, John A.
```

```
Query Match 99.8%; Score 1540; DB 9; Length 492;
Best Local Similarity 99.6%; Pred. No. 6.4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

/ 1 MKLNTSAGNVDIYKLYHSDACSSKAVYSLRCLACGVNLNSRSGRIYGGSSALPGAMPW 60
```

Db 210 MKLNTSAGNVDIYKKLYHSDACSKAVVSLRCLACGVNLSSRSRSHIVGESALPGAMPW 269
Qy 61 QVSLHVONVHVCSSGSIITPEWIVTAACHCEKPLNPNHMTAFAGILRQSFMYGAGYOVE 120
Db 270 QVSLHVONVHVCSSGSIITPEWIVTAACHCEKPLNPNHMTAFAGILRQSFMYGAGYOVE 329
Qy 121 KVISHPNDSTKKNNDIALMKLOKPLTFNDLVKPVCLPNPGMLOPBOUCWISGWGATEE 180
Db 330 KVISHPNDSTKKNNDIALMKLOKPLTFNDLVKPVCLPNPGMLOPBOUCWISGWGATEE 389
Qy 181 KKTSEVLANAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGGDSGGLVTSK 240
Db 390 KKTSEVLANAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGGDSGGLVTSK 449
Qy 241 NNIMWLIIGTSMGSCAKAYRPGYGVNVPFTDWTYRORADG 283
Db 450 NNIMWLIIGTSMGSCAKAYRPGYGVNVPFTDWTYRORADG 492

RESULT 12
US-09-895-793-932
Sequence 932, Application US/09895793
Publication No. US20020192763A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Devin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuguu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Jasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Baseola, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.534C2
CURRENT APPLICATION NUMBER: US/09/895,793
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 982
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-793-932

Query Match 99.8%; Score 1540; DB 10; Length 492;
Best Local Similarity 99.6%; Pred. No. 6.4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 330 KVISHPNDSTKKNNDIALMKLOKPLTFNDLVKPVCLPNPGMLOPBOUCWISGWGATEE 389
Qy 181 KKTSEVLANAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGGDSGGLVTSK 240
Db 390 KKTSEVLANAKVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGGDSGGLVTSK 449
Qy 241 NNIMWLIIGTSMGSCAKAYRPGYGVNVPFTDWTYRORADG 283
Db 450 NNIMWLIIGTSMGSCAKAYRPGYGVNVPFTDWTYRORADG 492

RESULT 13
US-09-895-814-932
Sequence 932, Application US/09895814
Publication No. US20020193296A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Devin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuguu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darlick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Jasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hurai, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Baseola, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C26
CURRENT APPLICATION NUMBER: US/09/895,814
CURRENT FILING DATE: 2001-06-29
NUMBER OF SEQ ID NOS: 990
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-895-814-932

Query Match 99.8%; Score 1540; DB 10; Length 492;
Best Local Similarity 99.6%; Pred. No. 6.4e-147;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 450 NNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWMYRQMRADG 492

RESULT 14

US-10-012-896-932

/ Sequence 932, Application US/10012896
/ Publication No. US20020183251A1
/ GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yugu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Derrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Baebols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
APPLICANT: Mantanabe, Yoshihiro
APPLICANT: Meagher, Madeleine Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C27
CURRENT APPLICATION NUMBER: US/10/012,896
CURRENT FILING DATE: 2001-12-10
NUMBER OF SEQ ID NOS: 1011
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 932
LENGTH: 492
TYPE: PRF
ORGANISM: Homo sapiens
US-10-012-896-932

Query Match 99.8%; Score 1540; DB 14; Length 492;

Best Local Similarity 99.6%; Pred. No. 6.4e-147;

Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNDVYKGLYHSDACSSKAVSLRCLACGVNINSSRSGRIYVGSBALPGAMPW 60
DB 210 MKLNTSAGNDVYKGLYHSDACSSKAVSLRCLACGVNINSSRSGRIYVGSBALPGAMPW 269
QY 61 QVSLHVQNVHVCSSITTPEMIVTAHCVKPLNPMWHTAFAGILROSFMFYGAGYOVE 120
DB 270 QVSLHVQNVHVCSSITTPEMIVTAHCVKPLNPMWHTAFAGILROSFMFYGAGYOVE 329
QY 121 KVISHPVDSKTKNDIALMKLOKPLTFNDLVYVCLPFGMWLQPEOLCWSIGMGATEE 180
DB 330 KVISHPVDSKTKNDIALMKLOKPLTFNDLVYVCLPFGMWLQPEOLCWSIGMGATEE 389
QY 181 KGMTSEVLANAKYLLITETORCNSRYYYDNLITPAMICAGFLQGVNDSQCGSGGGLVTSK 240
DB 390 KGMTSEVLANAKYLLITETORCNSRYYYDNLITPAMICAGFLQGVNDSQCGSGGGLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWMYRQMRADG 283
DB 450 NNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWMYRQMRADG 492

RESULT 15

US-10-205-823-415

/ Sequence 415, Application US/10205823

/ Publication No. US20030108963A1

/ GENERAL INFORMATION:

APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John B.
APPLICANT: Endege, Wilson O.
APPLICANT: Gannavarapu, Manjula
APPLICANT: Gorbacheva, Bella
APPLICANT: Hoersch, Sebastian
APPLICANT: Kamatkar, Shudhang
APPLICANT: Monsey, Angela W.
APPLICANT: Glatz, Karen
APPLICANT: Zhao, Xumei
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF PROSTATE CANCER
FILE OF INVENTION: THERAPY OF PROSTATE CANCER
FILE REFERENCE: MRI-044
CURRENT APPLICATION NUMBER: US/10/205,823
CURRENT FILING DATE: 2002-07-25
PRIOR APPLICATION NUMBER: 60/307,982
PRIOR FILING DATE: 2001-07-25
PRIOR APPLICATION NUMBER: 60/314,356
PRIOR FILING DATE: 2001-08-22/325,020
PRIOR APPLICATION NUMBER: 60/341,746
PRIOR FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/362,158
PRIOR FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 415
LENGTH: 492
TYPE: PRF
ORGANISM: Homo sapiens
US-10-205-823-415

Query Match 99.8%; Score 1540; DB 15; Length 492;

Best Local Similarity 99.6%; Pred. No. 6.4e-147;

Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNDVYKGLYHSDACSSKAVSLRCLACGVNINSSRSGRIYVGSBALPGAMPW 60
DB 210 MKLNTSAGNDVYKGLYHSDACSSKAVSLRCLACGVNINSSRSGRIYVGSBALPGAMPW 269
QY 61 QVSLHVQNVHVCSSITTPEMIVTAHCVKPLNPMWHTAFAGILROSFMFYGAGYOVE 120
DB 270 QVSLHVQNVHVCSSITTPEMIVTAHCVKPLNPMWHTAFAGILROSFMFYGAGYOVE 329
QY 121 KVISHPVDSKTKNDIALMKLOKPLTFNDLVYVCLPFGMWLQPEOLCWSIGMGATEE 180
DB 330 KVISHPVDSKTKNDIALMKLOKPLTFNDLVYVCLPFGMWLQPEOLCWSIGMGATEE 389
QY 181 KGMTSEVLANAKYLLITETORCNSRYYYDNLITPAMICAGFLQGVNDSQCGSGGGLVTSK 240
DB 390 KGMTSEVLANAKYLLITETORCNSRYYYDNLITPAMICAGFLQGVNDSQCGSGGGLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWMYRQMRADG 283
DB 450 NNIMWLIGDTSWGSCKAKAYRPGVGNVWFTDWMYRQMRADG 492

Search completed: August 1, 2003, 18:21:16

Job time : 54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2003, 18:10:24 ; Search time 18 Seconds

(without alignments)
1511.985 Million cell updates/sec

Title: US-09-988-975A-1

Perfect score: 1543

Sequence: 1 MKLNTSAGNVDIYKKLYHSD.....YYGNVMTFTDIYKQWRADG 283

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR76:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	555.5	36.0	638	1 KOHUP	plasma kallikrein
2	541	35.1	638	1 KOHUP	plasma kallikrein
3	536.5	34.8	1035	1 A53090	enteropeptidase (E
4	528.5	34.3	1019	1 A53090	enteropeptidase (E
5	526.5	34.1	638	1 KQMSPL	plasma kallikrein
6	520.5	33.7	417	1 S00845	hepsin (EC 3.4.21.
7	520	33.7	625	1 KFHU1	coagulation factor
8	509.5	33.0	855	2 JC7731	membrane-bound arg
9	505.5	32.8	812	1 PLMS	plasma (EC 3.4.21
10	503.5	32.6	416	1 S33777	hepsin (EC 3.4.21
11	502.5	32.6	1034	1 A53663	enteropeptidase (E
12	492	31.9	421	1 S11674	acrosin (EC 3.4.21
13	490	31.8	1524	2 T30337	polypeptidase - Afri
14	481.5	31.2	415	1 A34170	acrosin (EC 3.4.21
15	480.5	31.1	436	2 JX0172	acrosin (EC 3.4.21
16	467	30.3	418	2 A37344	acrosin (EC 3.4.21
17	466.5	30.2	431	2 S47538	acrosin (EC 3.4.21
18	464.5	30.1	421	2 S29599	acrosin (EC 3.4.21
19	461.5	29.9	437	2 S18407	acrosin (EC 3.4.21
20	459	29.7	1113	2 JEB315	low-density lipopr
21	458.5	29.7	790	1 PLRG	plasma (EC 3.4.21
22	454	29.4	810	1 A16260	plasma (EC 3.4.21
23	452	29.3	343	1 A57014	proctasin (EC 3.4.
24	449.5	29.1	810	2 B30848	plasma (EC 3.4.21
25	447.5	29.0	460	2 B61545	plasma (EC 3.4.21
26	444	28.8	812	1 PLBO	plasma (EC 3.4.21
27	441.5	28.6	267	2 S40006	trypsin (EC 3.4.21
28	441	28.6	270	2 S55160	maat cell trypsin
29	440	28.5	275	2 S40005	trypsin (EC 3.4.21

30	439	28.5	277	2 S35340	trypsin (EC 3.4.21
31	438	28.4	276	2 A38654	maat cell proteina
32	437.5	28.4	2	A55283	acrosin (EC 3.4.21
33	435.5	28.2	455	2 A61545	plasma (EC 3.4.21
34	435.5	28.2	810	1 PLHU	plasma (EC 3.4.21
35	434.5	28.2	266	2 S54146	trypsin (EC 3.4.21
36	433.5	28.1	4548	1 S00657	apoptoferrin(a) (EC
37	432.5	28.0	274	2 JC4171	trypsin (EC 3.4.2
38	430	27.9	273	2 A47246	trypsin (EC 3.4.2
39	429	27.8	274	2 S35339	trypsin (EC 3.4.21
40	428	27.7	275	2 S40007	trypsin (EC 3.4.21
41	427	27.7	242	2 S49489	trypsin (EC 3.4.21
42	425	27.5	263	2 A21195	chymotrypsin (EC 3
43	424	27.5	263	1 KYRFB	chymotrypsin (EC 3
44	424	27.5	275	2 C35863	trypsin (EC 3.4.2
45	423.5	27.4	263	2 A31299	chymotrypsin (EC 3

ALIGNMENTS

RESULT 1

KOHUP

N:Alternate names: Kininogenin, plasma prekallikrein

C:Species: Homo sapiens (man)

C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999

C:Accession: A00921; A37939

R:Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.

A:Title: Human plasma prekallikrein, a zymogen from a serine protease that contains four

A:Reference number: A00921; PMID:86243359; PMID:3521732

A:Accession: A00921

A:Molecule type: mRNA

A:Residues: 1-638 <CHU>

A:Cross-references: GB:M13143; NID:G190262; PIDN:AAA60153.1; PID:G190263

R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.

Biochemistry 30, 2050-2056, 1991

A:Title: Location of the disulfide bonds in human plasma prekallikrein: the presence c

A:Reference number: A37939; PMID:9152016; PMID:199866

A:Accession: A37939

A:Molecule type: protein

A:Residues: 20-27,40-46,'X',48,'H',50,'X',52-70,'H',75-76,'X',78-80,103-113,131-140,14

560-283,'X',285-287-291,'X',293-295,314-317,'X',319-320,331-324,'X',328-333,334-339,'

575-538-551,562,'X',564-567,573,'X',575-576,578-583,'X',585,592-604 <MCN>

C:Comment: This protein, synthesized by factor XIIIa, which cleaves the molecule into a

are linked by one or more disulfide bonds.

C:Comment: The zymogen is activated by factor XIIIa, which cleaves the molecule into a

inogen and may also play a role in the renin-angiotensin system by converting prorenin

A:Genetics: GDB:KXK3

A:Cross-references: GDB:127575; OMIM:229000

A:Map position: 4q35-4q35

C:Superfamily: coagulation factor XI; trypsin homology

C:Keywords: blood coagulation; duplication; fibrinolytic; glycoprotein; hydrolase; inf

F:1-19/Domin: signal sequence; duplication; fibrinolytic; glycoprotein; hydrolase; inf

F:20-638/Product: plasma kallikrein #status predicted <SIG>

F:20-390/Domin: plasma kallikrein heavy chain #status predicted <MAT>

F:110-109/Domin: apple repeat <AP1>

F:110-109/Domin: apple repeat <AP2>

F:200-289/Domin: apple repeat <AP3>

F:291-360/Domin: apple repeat <AP4>

F:391-621/Domin: plasma kallikrein light chain #status predicted <LCH>

F:391-621/Domin: trypsin homology <TRY>

F:121-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,3

F:127-308,396,453,494/Binding site: carboxylate (Asn) (covalent) #status experimental

F:318-347,340-345/Disulfide bonds: #status predicted

F:390-391/Cleavage site: Arg-11e (coagulation factor XIIIa) #status predicted

F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 36.0%; Score 555.5; DB 1; Length 638;

Best Local Similarity 41.8%; Pred. No. 2,4e-42;

C/Complex: mature enteropeptidase is variously reported to contain two (heavy and light) lifide linked

C/Function:

A/Description: cleaves propeptide from trypsinogen to produce active trypsin

A/Pathway: intestinal digestive hydrolase cascade

C/Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding reg

C/Keywords: glycoprotein; hydrolyase; serine proteinase; transmembrane protein

F:122-38/Domain: transmembrane #status predicted <TM>

F:52-117/Product: enteropeptidase mini chain #status predicted <HC>

F:118-800/Product: enteropeptidase heavy chain #status predicted <HC>

F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:358-520/Domain: MAM homology <MAM>

F:542-647/Domain: Clr/Cls repeat homology <Clr>

F:659-693/Domain: LDL receptor ligand-binding repeat homology #status atypical <SHC

F:801-1035/Domain: scavenger receptor cysteine-rich domain homology

F:801-1035/Domain: enteropeptidase light chain #status predicted <LCH>

F:116-147/170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binding

F:788-912,826-842,926-993,957-972,983-1011/Disulfide bonds: #status predicted

F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 34.8%; Score 536.5; DB 1; Length 1035;

Best Local Similarity 39.2%; Pred. No. 2,3e-40;

Matches 111; Conservative 53; Mismatches 104; Indels 15; Gaps 6;

Qy 1 MKNTSAGNDVYKYLHSDACSKRAVSLRC-IACGVNLSRQ-SRIYGSALPGA 57

Db 755 VNNTVTPNGSLI---LTPSQCLDLSLILLCNVYSCGKLVTCFVSFKYIGSGDSREGA 811

Qy 58 WPMOVSLHVNVAHVCGSITTPEMVTAACHVEKPLNPMWHTAFAGILRQSPFMYGAY 117

Db 812 WPMVVALYFPDOOVCAASLVSRMLVSAHCYGRNMRSKKAVLGLHMASNL---TSP 868

Qy 118 QVE---KVISHPNDSKTKNDIALMLQKLPFLFNDLVKVCPLNPMGLQPEOLCWI 172

Db 869 QITRLDQIVIPHNKRKNDIAMMLERKVVYTYIOTICLPENQVPPPRICSI 928

Qy 173 SGNGATEKSEKTESEVNAAKVLLIFTORCNSRYVDNITPAMICAGFLQGNVSCGDS 232

Db 929 AGMGALTYGSTADVLOEADVPLLSNERCQOCMPEN-ITEMMVCAGYRAGVSDSCGDS 987

Qy 233 GSKLVTSKNINIMLIDPTSGSGCAKAVRPGYGVNMFPTMI 275

Db 988 GGLPMCOENRMFLAGVTSFGYCALPMPRGVAVRPTETMI 1030

RESULT 4

A56318

enteropeptidase (EC 3.4.21.9) precursor [validated] - human

N/Alternate names: enterokinase

C/Species: Homo sapiens (man)

C/Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003

C/Accession: A56318; B43090

R/Kilamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.R.

Biochemistry 34, 4562-4568, 1995

A/Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic

A/Reference number: A56318; MIMD:95234679; PMID:7718557

A/Accession: A56318

A/Molecule type: mRNA

A/Residues: 1-1019 <KIT>

A/Cross-references: GB:009860; NID:g746412; PIDN:AA05318.1; PID:g746413

R/Kilamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A/Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo

A/Reference number: A43090; MIMD:94329561; PMID:8052624

A/Accession: B43090

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 749-1019 <KIT>

A/Cross-references: GB:1009860

C/Comment: The mechanism of association with the membrane of the intestinal brush border

ated below) or with amino-terminal myristoylation of the heavy chain.

C/Genetics:

A/Genes: GDB:PRSS7

A/Cross-references: GDB:384083; OMIM:226200

A/Map position: 21q21-21q21

C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig) ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv

ducts.

C/Function:

A/Description: cleaves activation peptide from trypsinogen to produce active trypsin

A/Pathway: intestinal digestive hydrolase cascade

C/Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding

C/Keywords: glycoprotein; hydrolyase; serine proteinase; transmembrane protein; zymoge

F:1-784/Product: enteropeptidase heavy chain #status predicted <HC>

F:22-38/Domain: transmembrane #status predicted <TM>

F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F:342-504/Domain: MAM homology <MAM>

F:526-631/Domain: Clr/Cls repeat homology <Clr>

F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <

F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>

F:785-1014/Domain: trypsin homology <TRY>

F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding sit

F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted

F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 34.3%; Score 528.5; DB 1; Length 1019;

Best Local Similarity 38.6%; Pred. No. 1.2e-39;

Matches 108; Conservative 56; Mismatches 107; Indels 9; Gaps 5;

Qy 1 MKNTSAGNDVYKYLHSDACSKRAVSLRC-IACGVNLSR-QSRIYGSALPGA 57

Db 739 VKNTAP---DGHILTPSQCLDLSLILLCNVYSCGKLVTCFVSFKYIGSGNAKGA 795

Qy 58 WPMOVSLHVNVAHVCGSITTPEMVTAACHVEKPLNPMWHTAFAGILRQSPFMYGAY 115

Db 796 WPMVVALYGGRLICGASLVSDMLVSAHCYGRNMRSKKAVLGLHMASNLTPQTV 855

Qy 116 GYVEKVIHPNDSKTKNDIALMLQKLPFLFNDLVKVCPLNPMGLQPEOLCWI 175

Db 856 PRILDSIVIPHNKRKNDIAMMLERKVVYTYIOTICLPENQVPPPRICSI 915

Qy 176 GATEKSEKTESEVNAAKVLLIFTORCNSRYVDNITPAMICAGFLQGNVSCGDSGX 235

Db 916 GTVVYCGTANIILOEADVPLLSNERCQOCMPEN-ITEMMVCAGYRAGVSDSCGSP 974

Qy 236 LVTSKNINIMLIDPTSGSGCAKAVRPGYGVNMFPTMI 275

Db 975 LMCQENRMFLAGVTSFGYCALPMPRGVAVRPTETMI 1014

RESULT 5

K065PL

plasma kallikrein (EC 3.4.21.34) precursor - mouse

C/Species: Mus musculus (house mouse)

C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999

C/Accession: A6557

R/Seldah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaudien, G.; Brachpapa, L.; Rocher

DNA Cell Biol. 9, 737-748, 1990

A/Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and compari

A/Reference number: A6557; MIMD:91090844; PMID:2264928

A/Accession: A6557

A/Molecule type: mRNA

A/Residues: 1-638 <SRI>

A/Cross-references: GB:458588; NID:g200358; PIDN:AAA6393.1; PID:g200359

A/Note: part of this sequence, including the amino ends of both the heavy and light ch

C/Comment: This protein, synthesized in the liver, circulates as a noncovalent complex

are linked by one or more disulfide bonds.

C/Superfamily: coagulation factor XI; trypsin homology

C/Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolyase; int

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>

F:20-109/Domain: apple repeat <AP1>

F:110-199/Domain: apple repeat <AP2>

```
F/200-289/Domain: apple repeat <AP3>
F/251-380/Domain: apple repeat <AP4>
F/331-638/Product: Plasma kallikrein light chain #status experimental <LCH>
F/331-621/Product: trypsin homology <TRY>
F/212-104-47-77-51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322-347,368-399,401-424,431-454,461-484,491-514,521-544,551-574,581-604,611-634,641-664,671-694,701-724,731-754,761-784,791-814,821-844,851-874,881-904,911-934,941-964,971-994,1001-1024,1031-1054,1061-1084,1091-1114,1121-1144,1151-1174,1181-1204,1211-1234,1241-1264,1271-1294,1301-1324,1331-1354,1361-1384,1391-1414,1421-1444,1451-1474,1481-1504,1511-1534,1541-1564,1571-1594,1601-1624,1631-1654,1661-1684,1691-1714,1721-1744,1751-1774,1781-1804,1811-1834,1841-1864,1871-1894,1901-1924,1931-1954,1961-1984,1991-2014,2021-2044,2051-2074,2081-2104,2111-2134,2141-2164,2171-2194,2201-2224,2231-2254,2261-2284,2291-2314,2321-2344,2351-2374,2381-2404,2411-2434,2441-2464,2471-2494,2501-2524,2531-2554,2561-2584,2591-2614,2621-2644,2651-2674,2681-2704,2711-2734,2741-2764,2771-2794,2801-2824,2831-2854,2861-2884,2891-2914,2921-2944,2951-2974,2981-3004,3011-3034,3041-3064,3071-3094,3101-3124,3131-3154,3161-3184,3191-3214,3221-3244,3251-3274,3281-3304,3311-3334,3341-3364,3371-3394,3401-3424,3431-3454,3461-3484,3491-3514,3521-3544,3551-3574,3581-3604,3611-3634,3641-3664,3671-3694,3701-3724,3731-3754,3761-3784,3791-3814,3821-3844,3851-3874,3881-3904,3911-3934,3941-3964,3971-3994,4001-4024,4031-4054,4061-4084,4091-4114,4121-4144,4151-4174,4181-4204,4211-4234,4241-4264,4271-4294,4301-4324,4331-4354,4361-4384,4391-4414,4421-4444,4451-4474,4481-4504,4511-4534,4541-4564,4571-4594,4601-4624,4631-4654,4661-4684,4691-4714,4721-4744,4751-4774,4781-4804,4811-4834,4841-4864,4871-4894,4901-4924,4931-4954,4961-4984,4991-5014,5021-5044,5051-5074,5081-5104,5111-5134,5141-5164,5171-5194,5201-5224,5231-5254,5261-5284,5291-5314,5321-5344,5351-5374,5381-5404,5411-5434,5441-5464,5471-5494,5501-5524,5531-5554,5561-5584,5591-5614,5621-5644,5651-5674,5681-5704,5711-5734,5741-5764,5771-5794,5801-5824,5831-5854,5861-5884,5891-5914,5921-5944,5951-5974,5981-6004,6011-6034,6041-6064,6071-6094,6101-6124,6131-6154,6161-6184,6191-6214,6221-6244,6251-6274,6281-6304,6311-6334,6341-6364,6371-6394,6401-6424,6431-6454,6461-6484,6491-6514,6521-6544,6551-6574,6581-6604,6611-6634,6641-6664,6671-6694,6701-6724,6731-6754,6761-6784,6791-6814,6821-6844,6851-6874,6881-6904,6911-6934,6941-6964,6971-6994,7001-7024,7031-7054,7061-7084,7091-7114,7121-7144,7151-7174,7181-7204,7211-7234,7241-7264,7271-7294,7301-7324,7331-7354,7361-7384,7391-7414,7421-7444,7451-7474,7481-7504,7511-7534,7541-7564,7571-7594,7601-7624,7631-7654,7661-7684,7691-7714,7721-7744,7751-7774,7781-7804,7811-7834,7841-7864,7871-7894,7901-7924,7931-7954,7961-7984,7991-8014,8021-8044,8051-8074,8081-8104,8111-8134,8141-8164,8171-8194,8201-8224,8231-8254,8261-8284,8291-8314,8321-8344,8351-8374,8381-8404,8411-8434,8441-8464,8471-8494,8501-8524,8531-8554,8561-8584,8591-8614,8621-8644,8651-8674,8681-8704,8691-8714,8721-8744,8751-8774,8781-8804,8811-8834,8841-8864,8871-8894,8901-8924,8931-8954,8961-8984,8991-9014,9021-9044,9051-9074,9081-9104,9111-9134,9141-9164,9171-9194,9201-9224,9231-9254,9261-9284,9291-9314,9321-9344,9351-9374,9381-9404,9411-9434,9441-9464,9471-9494,9501-9524,9531-9554,9561-9584,9591-9614,9621-9644,9651-9674,9681-9704,9691-9714,9721-9744,9751-9774,9781-9804,9811-9834,9841-9864,9871-9894,9901-9924,9931-9954,9961-9984,9991-10014,10021-10044,10051-10074,10081-10104,10111-10134,10141-10164,10171-10194,10201-10224,10231-10254,10261-10284,10291-10314,10321-10344,10351-10374,10381-10404,10411-10434,10441-10464,10471-10494,10501-10524,10531-10554,10561-10584,10591-10614,10621-10644,10651-10674,10681-10704,10691-10714,10721-10744,10751-10774,10781-10804,10811-10834,10841-10864,10871-10894,10901-10924,10931-10954,10961-10984,10991-11014,11021-11044,11051-11074,11081-11104,11111-11134,11141-11164,11171-11194,11201-11224,11231-11254,11261-11284,11291-11314,11321-11344,11351-11374,11381-11404,11411-11434,11441-11464,11471-11494,11501-11524,11531-11554,11561-11584,11591-11614,11621-11644,11651-11674,11681-11704,11691-11714,11721-11744,11751-117
```

Query Match	Score	DB	Length
Best Local Similarity	39.1%	Pred. NO. 16-39;	
Matched 105	45	W/matched 95	Toldo 73
			Can 5

QY 15 KLYHSPACSKAVVLSRCLACGVNINSKROSTIVGSEFALPGAMPQVSLHGVN--HY 71
Db 376 KLVDSBDCTTK-----NRIYGVNALSIGRFMQVSLQVLYVQSTH 418
QY 72 CGGSITTPEMVITAAHCEKPELNNPMTAFAGILRQSPFYGA-GYQVEKVISHPNYDS 130
Db 419 CGGSIIGRQVWLVAALHAFCD-GIPYDVMRIYGGILSLBIKMTPESSRIKELLTHOREYK 477
QY 131 KTKANNIALMKLOKPEFTENDLVKPYCLPMPGMLQPEQLCWSGSGATBEKGKTSFYNA 190
Db 478 SEGNYDIALIKLOFLPNTYTFPOKEPLCSKADNTNTIYNCAWYGTGKTEGSEFQNIQK 537
QY 191 AKVLLIEFQRCNSRYVDNLITPAMIAGPFLQGNVDSQSDSGSLVTSKNNIWMILIGDT 250
Db 538 ATTPLVANECCOKRY-RDYVINKQMI CAGYEKCGTDACKGSGGPIVCKSHSGEQLVGIT 596
QY 251 SMGSGAKAARPGVGNVAVNPFMDMYRQMA 281
Db 597 SMGEGGCKRQDPSVITKYSEYMDMILEKQDS 627

RESULT 6
S00845
hepsin (EC 3.4.21.-) - human
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C/Accession: S00845
R/Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
Biochemistry 27, 1067-1074, 1988
A/Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain
A/Reference number: S00845; PMID:88209431; PMID:2835076
A/Accession: S00845

```

A:Residues: 1-417 <LEV>
A:Cross-references: EMBL:X07732; NID:g32063; PDB:CAA3056.1; PDB:g32064
C:Genetics:
A:Gene: GDB:HEN; TMRSS1; hepsin
A:Cross-references: GDB:135685; OMIM:142440
A:Map position: 19q11-19q13.2
C:Superfamily: hepsin; trypsin homology
C:Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F:23-45/Domain: transmembrane #status predicted <TRN>
F:163-400/Domain: trypsin homology <TRY>
F:188-204,291-359,332-338,349-381/Disulfide bonds: #status predicted
F:203,257,353/Active site: His, Asp, Ser #status predicted

Query Match      33.7%  Score 520.5; DB 1; Length 417;
Best Local Similarity 38.7%; Pred. No. 2,1e-39;
Matches 104; Conservative 43; Mismatches 109; Indels 13; Gaps 4;

QY 22 CSSKAVYSURCIAQGVNLTSSRSRSTVGGESESLPGAMWQVSLVGVVCGSSTTPPM 81
D 140 CPKGRFLPALICDGG--RRKLVPDKIVGGRTSLGRMPQVSLRYDGAHLGGSLISGDM 197
QY 82 IVTPAHCEKPELNNPWHMTAPAGILRQSEFMFYGAGYQYEVKVISHPN-----DSKTKN 135
D 198 VLTPLACFPERRNRYLSRMWRFPAGAVAAQSP--HGLQLGQAVLVVYGGYLPRLPDNSSENEN 256
QY 136 DIALMKIQKELFTNDILKVPVCLPNPGMMIQPEOLQWISGMGATSEKTESEVYNAKATLL 195
D 257 DIALVHLSSELPLETKIYQPVCLPAAGQALVDGKLTCTVAGKGNATQYTGQAQGVQARVPI 316

```

Qy	196	IEFORCNSRYVDNLITTPMII	CAGTCLQGVNS	CGDSGGLY	-----TSKNIMWLI	LDTS	251
Db	317	ISNDVGNAGDFGNQIKRPMFC	CAGYEGSIDACQ	CGDSGSPV	CEDSISRTPRM	LCGIS	376
Qy	252	WGSGCAKAYRPGYGVNVWF	DDIVRQR				280
Db	377	WGTCALADQPGVYTKVSD	FRFMITQALIK				405

RESULT 7
KPH01

Query Match	33.7%;	Score 520;	DB 1;	Length 625;
Best Local Similarity	41.8%;	Pred. No. 3.9e-39;		

Matches 102; Conservative 43; Mismatches 85; Indels 14; Gaps 5;

QY 46 RIYGGSSALPGAMPQVSLHVN--HYCGSSITTPETVTAHCEKPLINFWHTAF 102
 DB 387 RIYGGSSALPGAMPQVSLHVN--HYCGSSITTPETVTAHCEKPLINFWHTAF 102
 QY 103 AGILROSF-----FYGAGQYERKVIHPNYSKTKNDIALMLQELTFENDLVKPVCL 157
 DB 446 SGLINQSEIKEDTSFG-----VQELIHQYQMASSGDIALMLQELTFENDLVKPVCL 501
 QY 158 PNRGMALQEPQLCWISGKATBEKGTSEVLAANAVLLIFQRNSRYVDNLITPMIC 217
 DB 502 PSKGDNRVLYTDCWYTWGYSKRLDKIOWTLOKATPIVTEBCKRY-RGHKITHMIC 560
 QY 218 AGFLQGNVDSGCGSLVTSKNNIMWLIQPTSGSCAAYRPGYGVNVPFTDKIR 277
 DB 561 AGREGKACKDGSGPLSCGHNVMLVGLTSGECAQREPRGVITNVEYVDILE 620
 QY 278 QMRA 281
 DB 621 KTOA 624

RESULT 8

JC7731
 Membrane-bound arginine-specific serine proteinase precursor - rat
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 03-Feb-2003
 C/Accession: J07731, J07775
 R/Kishi, K.; Yamazaki, K.; Yasuda, I.; Yanagi, N.; Ichinose, M.; Tsuchiya, Y.; Athuda, J. Biochem. 130, 425-430, 2001
 A/Biochem. 130, 425-430, 2001
 A/Title: Characterization of a membrane-bound arginine-specific serine protease from rat
 A/Reference number: J07731, MUID:21421307; PMID:1153019
 A/Accession: J07731
 A/Molecule type: mRNA
 A/Residues: 1-855 <KIS>
 A/Cross-references: DDBJ:AB049189
 A/Experimental source: strain Male, 7-week-old
 R/Satomi, S.; Yamasaki, Y.; Tsunuki, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T. Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
 A/Title: A role for membrane-type serine protease (MT-SPI) in intestinal epithelial turn
 A/Reference number: J07731; PMID:11573963
 A/Accession: J07731
 A/Molecule type: mRNA
 A/Residues: 1-855 <SAT>
 A/Cross-references: DDBJ:AB037898
 C/Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease of specific proteins or peptides on the brushborder membranes. It also participates in lial migration and/or cell loss.
 C/Genetics:
 A/Gene: mt-spl
 A/Map position: basolateral cell surface
 C/Superfamily: membrane-bound arginine-specific serine proteinase
 C/Keywords: protein digestion

Query Match 33.0%; Score 509.5; DB 2; Length 855;
 Best Local Similarity 38.6%; Pred. No. 5e-38;
 Matches 108; Conservative 56; Mismatches 97; Indels 19; Gaps 9;

QY 8 GNVDIYKLYHSDACSSKAVVSLRACIAGVNLSSRSRIYVGSALPGAMPQVSLH-V 66
 DB 583 GNPEDDKKKDSDSDER-----NC-DQGLR-STRKARVVGNNADSGMPQVSLH-V 635
 QY 67 QNVHVGSSITTPETVTAHCEKPLINFWHTAF 102
 DB 636 GQGHICGASLSPDLVSAHCFQDETIFKYSIDHTMTATLGLIDQS-KRASAGVQENHL 694
 QY 120 EKVISHPNVSKTKNDIALMLQELTFENDLVKPVCL 157
 DB 695 KRIITHPSFNDPTGYDIALLEKRPAYSTVAVPDLDPNTHVFRAGKAIWVGHTK 754
 QY 180 EKGKISVLANAKVLLIFQRNSRYVDNLITPMICAGFLQGNVDSGCGSLVTS- 238

DB 755 EGGTALILKGRIRVNTQTCR--ILPQITPRMVCGLSGVDSGQDSDGLSSV 812
 QY 239 SKNNIMWLIQPTSGSCAAYRPGYGVNVPFTDKIR 278
 DB 813 EXDGRIFQAGVSMGSCAQNRNKGVTTRIPVRWIMQ 852

RESULT 9

PLMS
 plasmin (EC 3.4.21.7) precursor - mouse
 N/Contains: angiotatin; plasminogen
 C/Species: Mus musculus (house mouse)
 C/Date: 20-Sep-1991 #sequence_revision 01-Nov-1996 #text_change 18-Jun-1999
 C/Accession: A38514, S48202, S48203
 R/Degen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W. Genomics 8, 49-61, 1990
 A/Title: Characterization of the cDNA coding for mouse plasminogen and localization of
 A/Reference number: A38514; MUID:91184812; PMID:2081600
 A/Accession: A38514
 A/Molecule type: mRNA
 A/Residues: 1-812 <DRG>
 A/Cross-references: GB:U04766; MUID:9200402; PID:AA50168.1; PID:9200403
 R/Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D. Eur. J. Biochem. 224, 863-871, 1994
 A/Title: Characterization of the murine plasma fibrinolytic system.
 A/Reference number: S48202; MUID:95010076; PMID:752120
 A/Accession: S48202
 A/Molecule type: Protein
 A/Residues: 20-25 <LID>
 A/Accession: S48203
 A/Molecule type: protein
 A/Residues: 22-27 <LID>
 C/Comment: Plasminogen is synthesized by the kidney and is present in plasma and many
 C/Comment: plasminogen is converted into plasmin by plasminogen activators, both plaser
 C/Comment: immediately after dissociation from the clot. In the presence of the inhibitor, the acti
 C/Comment: e inhibitor, the activation involves also removal of the activation peptide.
 C/Comment: Stromelysin I (see PIR:KCMSS1) acts on plasminogen to produce angiotatin.
 C/Function:
 A/Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a
 A/Description: as the walls of the graafian follicle; also activates the urokinase-type plasminogen a
 A/Pathway: fibrinolysis
 C/Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homolo
 C/Keywords: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein; hy
 F/1-96/Domain: plasminogen-related protein precursor homology <PLPH>
 F/1-19/Domain: signal sequence #status predicted <SIG>
 F/20-812/Product: plasminogen #status predicted <PRO>
 F/20-96/Domain: activation peptide #status predicted <APT>
 F/79-466/Product: angiotatin #status predicted <AST>
 F/97-581/Domain: chain A #status predicted <A>
 F/103-181/Domain: kringe homology <KR1>
 F/185-262/Domain: kringe homology <KR2>
 F/275-352/Domain: kringe homology <KR3>
 F/377-454/Domain: kringe homology <KR4>
 F/481-560/Domain: kringe homology <KR5>
 F/582-812/Domain: chain B #status predicted <BCH>
 F/582-805/Domain: trypsin homology <TRY>
 F/49-73, 53-61, 103-181, 124-164, 152-176, 185-262, 188-316, 206-245, 234-257, 275-352, 296-335,
 bonds: #status predicted
 F/78-79/Cleavage site: Glu-Aen (stromelysin 1) #status predicted
 F/136, 308/Cleavage site: carboxylate (Asn) (covalent) #status predicted
 F/466-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted
 F/581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental
 F/624, 667, 762/Active site: His, Asp, Ser #status predicted

Query Match 32.8%; Score 505.5; DB 1; Length 812;
 Best Local Similarity 40.5%; Pred. No. 1.1e-37;
 Matches 113; Conservative 39; Mismatches 98; Indels 29; Gaps 9;

QY 14 KKLTHSDACSSKAVVSLRACIAGVNLSSRS--SRIYVGSALPGAMPQVSLHVN--V 69
 DB 813 EXDGRIFQAGVSMGSCAQNRNKGVTTRIPVRWIMQ 852

```

Db      549 RRLY--DYCDIPLCASASSFECKPQVEPKCPGRVVGCVANSHSNPWOISLRTEFTQ 606
Qy      70 HVGGSITTPEMIVTAHCEKPLNNPMTAFAGILROSEMFYAGGYE-----KVIS 124
      607 HFCGGTLAEPVLTAAHCEKLS--SRPEFYKVLGAHNE-----YIRLDVQETISVAKLL 661
Db      125 HENYSKTRNDIATMLKQPLTFNDLVKPYCLPMPQMLQPBELCMISQKTE---EK 181
Qy      662 EPN-----NRDIALKLKSRPATYTDKTPACPLSPMYVADRTTCYTTEMGETQTFGA 715
Db      182 GKTSEVLAHAKVLLIETORCNSRYVDNLITPAMICAGFLQGVNDSQCGDSGGKLVTSKN 241
Qy      716 GRLKE---AQPLVENKVCNRVEYLNRRVSTELCGQLAGVDSQCGDSGGKLVCPFK 771
Db      242 NIMWLLIGTSMGSGCAKAYRPGYKGNVWVFTDNYROR 280
Qy      772 DKYILQGVTSWELGCAKARPKGVYVRSRPFWMIRER 810

RESULT 10
hepsin (EC 3.4.21.-) - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999
C/Accession: S33777; S32013
R/Exley, D.; Raymond, F.; Nick, H.
Biochem. Biophys. Acta 1173, 350-352, 1993
A/Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
A/Reference number: S33777; MUID:93305733; PMID:8318546
A/Accession: S33777
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-416 <PRA>
A/Cross-references: EMBL:X70900; NID:G57928; PIDN:CAA50256.1; PID:G57929
C/Superfamily: hepsin; trypsin homology
C/Keyword: hydrolase; liver; serine proteinase; transmembrane protein
F:22-44/Domain: transmembrane #status predicted <TM>
F:162-399/Domain: trypsin homology <TRY>
F:187-203,299-358,321-337,348-380/Disulfide bonds: #status predicted
F:202,256,352/Active site: His, Asp, Ser #status predicted

Query Match      32.6%; Score 503.5; DB 1; Length 416;
Best Local Similarity 37.9%; Pred. No. 7,5e-38;
Matches 102; Conservative 45; Mismatches 109; Indels 13; Gaps 4;

Qy      22 GSKAVSLKRCIACGVNLSRSRIVGSEALGAMPQVSLVQVNHVCGSITTPM 81
Db      139 CRRGRFLITCODG--RRKLPRDRIVGQDSSIGRMWQVSLKTDGTHLGGSLSDM 196
Qy      82 IYTAHCEKPLNNPMTAFAGILROSEMFYAGGYEYVISHPNY-----DSKTKN 135
Db      197 VLTAAHCEKPLNNPMTAFAGILROSEMFYAGGYEYVISHPNY-----DSKTKN 255
Qy      136 DIALMLKQPLTFNDLVKPYCLPMPQMLQPBELCMISQKTESEVLAHAKVLL 195
Db      256 DIALMLKQPLTFNDLVKPYCLPMPQMLQPBELCMISQKTESEVLAHAKVLL 315
Qy      196 IETORCNSRYVDNLITPAMICAGFLQGVNDSQCGDSGGKLV-----TSKNINWLLIGTS 251
Db      316 ISENVCSNDFYGNQIKPMPFCAGYPEGIDACQDSGHVFCEDRISGSRWLGIYS 375
Qy      252 WSGGCAKAYRPGYKGNVWVFTDNYROR 280
Db      376 WGTGALAKKPGVYKVLDFRMTQALK 404

RESULT 11
enteropeptidase (EC 3.4.21.9) precursor [validated] - pig
N/Alternate names: enterokinase
C/Species: Sus scrofa domestica (domestic pig)
C/Date: 07-Oct-1994 #sequence_revision 09-Aug-1996 #text_change 28-Apr-2003
C/Accession: A53663

```

```

R/Matsushima, M.; Ichinohe, M.; Yahagi, N.; Kakei, N.; Takeda, S.; Miki, K.; Kurokawa
J. Biol. Chem. 269, 19976-19982, 1994
A/Title: Structural characterization of porcine enteropeptidase.
A/Reference number: A53663; MUID:94327548; PMID:8051081
A/Accession: A53663
A/Molecule type: mRNA
A/Residues: 1-1034 <MAT>
A/Cross-references: GB:ID30799; NID:G505122; PIDN:BA06459.1; PID:G505123
A/Note: parts of this sequence, including the amino ends of three chains isolated from
C/Comment: The mechanism of association with the membrane of the intestinal brush bord
C/Comment: (see below) or with amino-terminal myristoylation of the heavy chain.
C/Complex: Mature enteropeptidase is variously reported to contain two (heavy and ligh
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms invo
C/Function:
A/Description: cleaves activation peptide from trypsinogen to produce active trypsin
A/Pathway: intestinal digestive hydrolase cascade
C/Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding r
C/Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-799/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:357-519/Domain: MM homology <MM>
F:541-646/Domain: Clr/Cls repeat homology <CLR>
F:658-692/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <S
F:800-1034/Product: enteropeptidase light chain #status predicted <LCH>
F:116-147/Domain: trypsin homology <TRY>
F:116-147,170-194,283-343,350-403,455-485,518-549,645-697,701-721,740-761,804,863,902,
F:787-911,925-992,995-971,982-1010/Disulfide bonds: #status predicted
F:840,881,986/Active site: His, Asp, Ser #status predicted

Query Match      32.6%; Score 502.5; DB 1; Length 1034;
Best Local Similarity 36.4%; Pred. No. 2,7e-37;
Matches 103; Conservative 60; Mismatches 105; Indels 15; Gaps 6;

Qy      1 MKLNTSGNVGVYKYLHSDCSSKAVYLR--IACGVNLSRSR--RIVGSEALPG 57
Db      754 VKLNTSPNSL--LTSBQCFEBSLILCNHKSQKQVQVSPKIVGGDSREG 810
Qy      58 WPNQVSLHVNQVHVGGSITTPEMIVTAHCEKPLNNPMTAFAGILROSEMFYAGY 117
Db      117 WPNVVALYVNGQLTGASIVSRDMLVSAHCVYGRNLEPSKRAILGLHMTSNL--TSP 867
Qy      118 Q-----VEKVISHPYDSKTRNDIATMLKQPLTFNDLVKPYCLPMPQMLQPBELCM 172
Db      868 QIVTRLIDIVINPHNRRKQSDIAMWLEFFKVNITDYIQLPLPENQVPPRGICSI 927
Qy      173 SGWATEERKTSSEVLAHAKVLLIETORCNSRYVDNLITPAMICAGFLQGVNDSQCGDS 232
Db      928 AGGKVIYQGSADILQEDVPLLSNEKQCGQMEYV--ITENMWCAQYEGGIDSCGDS 986
Qy      233 GSKLVTSKNINWLLIGTSWGGCAKAYRPGYKGNVWVFTDNYROR 275
Db      987 GSKLVTSKNINWLLIGTSWGGCAKAYRPGYKGNVWVFTDNYROR 1029

RESULT 12
acrosin (EC 3.4.21.10) precursor - human
C/Species: Homo sapiens (man)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C/Accession: S11674; S23499; S12063; A61022; S03330
R/Kelme, S.; Adham, I.M.; Engel, W.
Eur. J. Biochem. 190, 195-200, 1990
A/Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene
A/Reference number: S11674; MUID:90306003; PMID:2114285
A/Accession: S11674
A/Molecule type: DNA
A/Residues: 1-421 <KEI>
A/Cross-references: EMBL:154017; NID:G35582; PIDN:CAA37964.1; PID:G1226165
A/Note: The authors translated the codon AGG for residue 64 as Thr and CTG for residue
R/Vazquez-Levin, M.H.; Reventos, J.; Gordon, J.W.

```


Eur. J. Biochem. 207, 23-26, 1992
 A>Title: Molecular cloning, sequencing and restriction mapping of the genomic sequence
 A:Reference number: S23499; MUID:9231659; PMID:1628652
 A:Accession: S23499
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-421 <VAZ>
 A:Cross-references: EMBL:M77378
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992.
 R:Keime, S.
 submitted to the EMBL Data Library, December 1989
 A:Reference number: S12063
 A:Accession: S12063
 A:Molecule type: DNA
 A:Residues: 1-225, 'R', 227-421 <KEI2>
 A:Cross-references: EMBL:X54017
 R:Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
 Hum. Genet. 84, 125-128, 1990
 A>Title: Molecular cloning of human preproacrosin cDNA.
 A:Reference number: A61022; MUID:90128988; PMID:2298447
 A:Accession: A61022
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>
 R:Baba, T.; Matanabe, K.; Kashiwabara, S.I.; Arai, Y.
 FEBS Lett. 244, 296-300, 1989
 A>Title: Primary structure of human proacrosin deduced from its cDNA sequence.
 A:Reference number: S03330; MUID:89153568; PMID:2493394
 A:Accession: S03330
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
 A:Cross-references: EMBL:Y00970; MUID:928325; PDB:CAA68784.1; PDB:928326
 C:Genetic:
 A:Gene: GDB:ACR
 A:Cross-references: GDB:119645; OMIM:102480
 A:Map position: 22q13-22qter
 A:Intons: 26/2; 94/2; 189/2; 237/3
 C:Superfamily: acrosin; trypsin homology
 C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-421/Product: acrosin #status predicted <MAI>
 F:20-42/Product: acrosin light chain #status predicted <LCH>
 F:43-421/Product: acrosin heavy chain #status predicted <HCH>
 F:43-285/Domain: trypsin homology <TRY>
 F:302-379/Region: proline-rich
 F:22-210/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:25-154/Disulfide bonds: #status predicted
 F:73-89/Disulfide bonds: #status predicted
 F:88-112, 240/Active site: His, Asp, Ser #status predicted
 F:117-246/Disulfide bonds: #status predicted
 F:209-225/Disulfide bonds: #status predicted
 F:236-266/Disulfide bonds: #status predicted
 Query Match 31.8%; Score 492; DB 1; Length 421;
 Best Local Similarity 40.5%; Pred. No. 8, 4e-37;
 Matches 106; Conservative 42; Mismatches 88; Indels 26; Gaps 10;
 QY 35 CGVNLNLSHQ--RIVGESALPGAMPVQVSLV-----QNVHVCSSITTEPIVYAAH 87
 DB 29 CGIRFQNPQGGRIIVGKRAQHGAMPVMSLQIRVYNSHRVYHCGSSLNSRVVLTAAH 88
 QY 88 C-VEKLNPNMHTAFAGILRSFMYG-----AGYV--YEVVISHPNYDSKTKNDI 137
 DB 89 CTVGK--NNVHDPRLVFGAKS---ITYGNNKRVKAPLQERVYKTIITHERYNSATGNDI 143
 QY 138 ALMKLQKPLTFNDLVKPVCLPN--PQMLQPSQLCISGSGATEBK--KTSEVLAIAKVL 195
 DB 144 ALVEIRPPIISCRFIPGCLPHFKAGLPRGQS CVVAGMGTIEKAPRPSILMEARVLD 203
 QY 196 IETORNSYVYVYDNITPAMICAGFLGQNVNCSGGDSGXV--TSKNIMWMLIGTSG 253
 DB 204 IDLDLNSGQWNGRWNGVQPIVNGAGIPVQKIDPDCGDSGFLMKDSKESATVAVGITSWG 263

QY 254 SGCAKAVRPGVYGVNMTFTMI 275
 DB 264 VGCLAKRPGIYATMPYLMVI 285
 RESULT 13
 T30337
 polypeptide - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 03-Feb-2003
 C:Accession: T30337
 R:Yang, J.C.; Lindsey, L.L.; Hedrick, J.L.
 submitted to the EMBL Data Library, March 1998
 A>Description: cDNA cloning of oocyte-maturation, a chymotrypsin-like protease released from
 A:Reference number: Z20829
 A:Accession: T30337
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1524 <YAN>
 A:Cross-references: EMBL:U01290; MUID:92981640; PDB:92981641; PDB:AA24717.1
 C:Superfamily: trypsin related polypeptide; trypsin homology
 Query Match 31.8%; Score 490; DB 2; Length 1524;
 Best Local Similarity 41.7%; Pred. No. 5, 9e-36;
 Matches 98; Conservative 43; Mismatches 84; Indels 10; Gaps 6;
 QY 46 RIVGESALPGAMPVQVSLVQNVHVCSSITTEPIVYAAHCVKELNPNMHTAFAGI 105
 DB 563 RIVGESALPNSFWQVQVLPFLTFHCBALISPOLITAAHCIRAA--EESYTVVING- 639
 QY 106 LRQSPFYGAGYQVYKYS---HPNYSKTKNDIALMKLQKPLTFNDLVKPVCLPNRM 162
 DB 640 -DHRMLNSETBOIRNKTIRIRIDNYSSTYNDIALVLEPDLNDFVAPVCLPEPE 698
 QY 163 MLEPQLCWSGSGATEBKKTSEVLAIAKVLITQKNSYVYDNITPAMICAGTLO 222
 DB 639 VLTASVCSYVTGNTAEDQPLAGLOQLTILDSIICNTSY-YSGETTHMCCAGPSS 757
 QY 223 G-NVDSQGSQSGXV--TSKNIMWMLIGTSGSGCAKAVRPGVYGVNMTFTMI 275
 DB 758 SKEDACQDSGSPVLCQVKEQFSYGLVSGBGGRVSKGVYTKRLFTMI 812
 RESULT 14
 A34170
 acrosin (EC 3.4.21.10) precursor - pig
 N:Alternate names: 53K fucose-binding protein
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A34170; S08994; S02428; S04940; S16557; S02780; S106957; S12968
 R:Baba, T.; Kashiwabara, S.; Matanabe, K.; Itoh, H.; Michiawa, Y.; Kimura, K.; Takada
 J. Biol. Chem. 264, 11920-11927, 1989
 A>Title: Activation and maturation mechanisms of boar acrosin zymogen based on the ded
 A:Reference number: A34170; MUID:89308595; PMID:2745422
 A:Accession: A34170
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-415 <BAB>
 A:Cross-references: GB:J04950; MUID:9164702; PDB:AAA31131.1; PDB:9164703
 R:Cechova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V.
 Biol. Chem. Hoppe-Seyler 371, 317-323, 1990
 A>Title: Is spermogen a modified proacrosin? Isolation, purification, and partial ch
 A:Reference number: S08994; MUID:90253655; PMID:2111146
 A:Accession: S08994
 A:Molecule type: protein
 A:Residues: 'X', 16, 'X', 20-25, 'X', 27-32, 'X', 34-38, 'X', 40-50 <BC>
 R:Toepfer-Petersen, E.; Henschen, A.
 FEBS Lett. 226, 38-42, 1987
 A>Title: Acrosin shows zona and fucose binding, novel properties for a serine protease
 A:Reference number: S02428; MUID:88083633; PMID:3480243
 A:Accession: S02428
 A:Molecule type: protein

A/Residues: 17-32,40-55 <TOE>
R/Adham, I.M.; Klemm, U.; Møller, W.M.; Hoyer-Fender, S.; Tsaousidou, S.; Engel, W.
Eur. J. Biochem. 182, 563-568, 1989
A/Title: Molecular cloning of preproacrosin and analysis of its expression pattern in sp
A/Reference number: S04940; MUID:89925501; PMID:2502391
A/Accession: S04940
A/Molecule type: mRNA
A/Residues: 1-7,9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-393, 'GN', 396, 'LVE', 399-40
A/Cross-references: EMBL:X14844
A/Note: the authors translated the codon CCT for residue 240 as Ala, GCC for residue 264
R/Adham, I.M.
submitted to the EMBL Data Library, March 1989
A/Reference number: S16657
A/Accession: S16657
A/Molecule type: mRNA
A/Residues: 1-7,9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-398, 'XELL', <AD2>
A/Cross-references: EMBL:X14844; NID:91867; PIDN:CA432948.1; PID:91868
A/Note: the difference at the carboxyl end is due to a frameshift error
R/Baba, T.; Michikawa, Y.; Kawakura, K.; Arai, Y.
FEBS Lett. 244, 132-136, 1989
A/Title: Activation of boar proacrosin is effected by processing at both N- and C-termini
A/Reference number: S02780; MUID:89171246; PMID:2494060
A/Accession: S02780
A/Molecule type: protein
A/Residues: 17-69 <BA2>
R/Toepfer-Petersen, E.; Steinberger, M.; von Eschenbach, C.E.; Zucker, A.
FEBS Lett. 265, 51-54, 1990
A/Title: Zona pellucida-binding of boar sperm acrosin is associated with the N-terminal
A/Reference number: S10695; MUID:90306316; PMID:2365054
A/Accession: S10695
A/Molecule type: protein
A/Residues: 40-62 <TO2>
R/Toepfer-Petersen, E.; Calvete, J.; Schaefer, W.; Henschen, A.
FEBS Lett. 275, 139-142, 1990
A/Title: Complete localization of the disulfide bridges and glycosylation sites in boar
A/Reference number: S12968; MUID:91085546; PMID:2261983
A/Accession: S12968
A/Molecule type: protein
A/Residues: 17-29;34-66;68-91;94-121;123-166;171-184;190-207;209-216;219-228;231-245;246
C/Superfamily: acrosin; trypsin homology
C/Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-415/Product: acrosin #status experimental <MAT>
F:17-39/Product: acrosin light (A) chain #status experimental <LCH>
F:40-415/Product: acrosin heavy (B) chain #status experimental <HC>
F:40-283/Domain: trypsin homology <TRY>
F:300-374/Region: proline-rich (Aan) (covalent) #status experimental
F:19,208/Binding site: carbohydrate (Aan) (covalent) #status experimental
F:22-152,26-160,71-87,175-244,207-223,234-264/Disulfide bonds: #status experimental
F:86,140,238/Active site: His, Asp, Ser #status predicted

Query Match 31.2% Score 481.5; DB 1; Length 415;
Best Local Similarity 37.3%; Pred. No. 7,4e-36;
Matches 103; Conservative 41; Mismatches 79; Indels 53; Gaps 9;

QY 35 CGVNLSSRQ--RIYGGESALPGAMPQVSLV-----ONVHCGGSIITPBWVITAA 86
DB 26 CGLRFRQKLESQMRVYGGVSAEPGAMPVWSLQIFMHNRRYHTCGIILNHWVITAA 85
QY 87 HCVENPNNPMTAFAGILRSFMYGAG-----YQEKVISHPNYDSKTKNDI 137
DB 86 HCFDNK-KKQYDMRLVFGAQR---IEGRNKPVRBPQGRVYQKIVIHKKNVYTBGNDI 144
QY 125 HPNYDSKTKNDIALLKLOKPLTNDLVKPVCLP--NQGMALPQOLCWISGMANGEEKG 182
DB 129 HEKYVSGLEINDIALIKITPPVPCGPPIGPCLPDPFKAGPPRAV-QTCWYTGKGYKXKG 187
QY 183 -KTSEVLNAAKYLLETCRNSRYVYDMLITPAMICAGPLQGVNDSQGGSGGLV--TS 239
DB 188 PRISPLQGARVALIDIELCNSTRVYNGIRISTVNCAGYPRGKIDTCQDSGGPLMCRDR 247
QY 240 KNNIMLIGDTSNGSGCAKAYRPGYGNVMTDWT 275

DB 248 AENTPVVVGITSMGVCARAKRPGVYTSTWPIYLMWI 283

RESULT 15
JX0172
acrosin (BC 3.4.21.10) precursor form 3 - mouse
C/Spectra: Mus musculus (house mouse)
C/Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text change 20-Jun-2000
C/Accession: JX0172; U0138
R/Watanabe, K.; Baba, T.; Kaishiwabara, S.; Okamoto, A.; Arai, Y.
J. Biochem. 109, 828-833, 1991
A/Title: Structure and organization of the mouse acrosin gene.
A/Reference number: JX0172; MUID:92041732; PMID:1939002
A/Accession: JX0172
A/Molecule type: DNA
A/Residues: 1-436 <MAT>
A/Cross-references: GB:S66245; NID:9238706; PIDN:AA820293.1; PID:9238707
R/Kaishiwabara, S.; Baba, T.; Takada, M.; Watanabe, K.; Yano, Y.; Arai, Y.
J. Biochem. 108, 785-791, 1990
A/Title: Primary structure of mouse proacrosin deduced from the cDNA sequence and its
A/Reference number: JX0138; MUID:91185535; PMID:2127931
A/Accession: JX0138
A/Molecule type: mRNA
A/Residues: 4-436 <BAS>
A/Cross-references: GB:D00754; NID:9220322; PIDN:BA800651.1; PID:9220323
C/Comment: Acrosin is an acrosomal protease that plays an important role in the initia
e-binding activity.
C/Genetics:
A/Introns: 26/2; 95/2; 190/1; 238/3
C/Superfamily: acrosin; trypsin homology
C/Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-42/Product: acrosin light chain #status predicted <ALC>
F:43-321/Product: acrosin heavy chain #status predicted <HC>
F:43-286/Domain: trypsin homology <TRY>
F:22,211/Binding site: carbohydrate (Aan) (covalent) #status predicted
F:125-155,29-162,74-90,178-247,210-225,237-267/Disulfide bonds: #status predicted
F:89,141,241/Active site: His, Asp, Ser #status predicted
F:233/Binding site: substrate (Asp) #status predicted

Query Match 31.1% Score 480.5; DB 2; Length 436;
Best Local Similarity 38.3%; Pred. No. 9.6e-36;
Matches 101; Conservative 45; Mismatches 89; Indels 29; Gaps 9;

QY 35 CGVNL--NSRQSRIVYGGESALPGAMPQVSLV-----ONVHCGGSIITPBWVITAA 86
DB 29 CGLRFRQKLESQMRVYGGVSAEPGAMPVWSLQIFMHNRRYHTCGIILNHWVITAA 85
QY 87 HCVENPNNPMTAFAGILRSFMYGAG-----YQEKVISHPNYDSKTKNDI 137
DB 89 HCFDNK-KKQYDMRLVFGAQR---IEGRNKPVRBPQGRVYQKIVIHKKNVYTBGNDI 144
QY 138 ALMKLOKPLTNDLVKPVCLP--NQGMALPQOLCWISGMANGEEKG-KTSEVLNAAKY 193
DB 145 ALMKITPPVTCAGPIGPCLP--FKAGPPQIPIHTCVYTGKGYKXKAPRSPVLEAKY 202
QY 194 LLLETCRNSRYVYDMLITPAMICAGPLQGVNDSQGGSGGLVTSNR--NIMWLIGDTS 251
DB 203 DLIDLDLNCSTQWYNGIRISTVNCAGYPRGKIDTCQDSGGPLMCRDNDSPFVVGITS 262
QY 252 WSGSGCAKAYRPGYGNVMTDWT 275
DB 263 WSGVGCARAKRPGVYTATMDYDWT 286

Search completed: August 1, 2003, 18:13:05
Job time : 19 secs

PI Bandman O, Lai P;
 XX WPI; 2000-282523/24.
 DR N-PSDB; AAI12975.
 XX
 PT Polynucleotide encoding human prostate-associated protease useful for
 PT diagnosing and treating cancers, prostate disorders and
 PT gastrointestinal disorders
 PS Claim 1; Fig 1A-B; 27pp; English.
 XX
 CC This sequence represents human prostate-associated protease (HUPAP). cDNA
 CC encoding HUPAP was initially identified in a spinal cord cDNA library.
 CC the cDNA encoding this sequence representing a consensus of overlapping
 CC and/or extended nucleic acid sequences from spinal cord, prostate tumour
 CC and colon cDNA libraries. HUPAP is a serine protease with structural and
 CC functional homology with bovine enterokinase, human pancreatic kallikrein
 CC and African rat renal kallikrein, sharing 38% homology with bovine
 CC enterokinase. In addition, HUPAP is related to prostate-specific antigen
 CC (PSA), a kallikrein which is a highly sensitive marker for prostate
 CC cancer. HUPAP and nucleic acids encoding it are useful for the diagnosis,
 CC prevention and treatment of gastrointestinal disorders such as ulcerative
 CC colitis, pancreatitis, cancers, and prostatic disorders. HUPAP agonists
 CC may be used to treat gastrointestinal disorders, and HUPAP antagonists
 CC and inhibitors may be used to suppress excessive cell proliferation,
 CC which is of use in cancer therapy. HUPAP nucleic acids are also useful
 CC for generating hybridisation probes which may be used for mapping
 CC naturally occurring genomic sequences.
 XX
 SQ Sequence 283 AA;
 Query Match 99.9%; Score 1541; DB 21; Length 283;
 Best Local Similarity 100.0%; Pred. No. 3.7e-136;
 Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCIACGVNLNSSRSQRIVGESALPGAWPM 60
 Db 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCIACGVNLNSSRSQRIVGESALPGAWPM 60
 Oy 61 QVSLHQNVAHVCGGSIITPEWITVAACVEKPLNPNHMTAFAGILRQSMFYGAGQYE 120
 Db 61 QVSLHQNVAHVCGGSIITPEWITVAACVEKPLNPNHMTAFAGILRQSMFYGAGQYE 120
 Oy 121 KYISHPNYDSKTKNNDIALMKLOKPLTFNDLVKVCPLPFGMLQPEQLCMISGWGATEE 180
 Db 121 KYISHPNYDSKTKNNDIALMKLOKPLTFNDLVKVCPLPFGMLQPEQLCMISGWGATEE 180
 Oy 181 KGIETSEVINAARVLLIETORCNRSRYVDNLTTPAMICAGFLQGNVDSGQDSGGLVTSK 240
 Db 181 KGIETSEVINAARVLLIETORCNRSRYVDNLTTPAMICAGFLQGNVDSGQDSGGLVTSK 240
 Oy 241 NNIMWLIGDTSWGSACAKAYRPGYGNVMTFTDIYRQMBADG 283
 Db 241 NNIMWLIGDTSWGSACAKAYRPGYGNVMTFTDIYRQMBADG 283
 RESULT 2
 ID ABB78328 standard; Protein; 283 AA.
 XX ABB78328;
 AC
 XX
 DT 05-DEC-2002 (first entry)
 XX
 DE Amino acid sequence of human prostate-associated protease (HUPAP).
 XX
 KW Human; prostate-associated protease; HUPAP; prostatic disorder;
 KW prostate cancer; benign prostatic hyperplasia; gastrointestinal disorder;
 KW congenital enterokinase deficiency; pancreatitis; ulcerative colitis;
 KW cancer; enzyme.
 XX
 OS Homo sapiens.

PH Key Location/Qualifiers
 FT Modified-site 4
 FT /note= "potential N-glycosylation site"
 FT Misc-difference 235
 FT /note= "unspecified residue encoded by CNT"
 XX
 PN US2002119531-A1.
 XX
 PD 29-AUG-2002.
 XX
 PF 19-NOV-2001; 2001US-0988975.
 XX
 PR 27-FEB-1997; 97US-0807151.
 PR 07-JUN-2000; 2000US-0478957.
 XX
 PA (BAND/) BANDMAN O.
 PA (LALF/) LAL P G.
 PI Bandman O, Lai P;
 XX
 DR WPI; 2002-627171/67.
 DR N-PSDB; ABB72262.
 XX
 PT Novel human prostate-associated protease (HUPAP), useful for treating
 PT gastrointestinal disorders such as congenital enterokinase deficiency
 PT and for producing anti-HUPAP antibodies -
 XX
 PS Claim 1; Fig 1A-B; 27pp; English.
 XX
 CC The present sequence represents human prostate-associated protease
 CC (HUPAP). The antibody that specifically binds to HUPAP is useful in the
 CC treatment of a prostatic (e.g. prostate cancer and benign prostatic
 CC hyperplasia) or gastrointestinal disorder (e.g. congenital enterokinase
 CC deficiency). The HUPAP antagonist is useful for treating the prostate
 CC disorders described above. The HUPAP antagonist or antibody is also
 CC useful for treating pancreatitis, ulcerative colitis and cancers of the
 CC oesophagus, stomach, small intestine, large intestine or colon. HUPAP or
 CC its agonist is useful for treating the gastrointestinal disorders
 CC described above.
 XX
 SQ Sequence 283 AA;
 Query Match 99.9%; Score 1541; DB 23; Length 283;
 Best Local Similarity 100.0%; Pred. No. 3.7e-136;
 Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCIACGVNLNSSRSQRIVGESALPGAWPM 60
 Db 1 MKLNTSAGNVDIYKKLYHSDACSSKAVVSLRCIACGVNLNSSRSQRIVGESALPGAWPM 60
 Oy 61 QVSLHQNVAHVCGGSIITPEWITVAACVEKPLNPNHMTAFAGILRQSMFYGAGQYE 120
 Db 61 QVSLHQNVAHVCGGSIITPEWITVAACVEKPLNPNHMTAFAGILRQSMFYGAGQYE 120
 Oy 121 KYISHPNYDSKTKNNDIALMKLOKPLTFNDLVKVCPLPFGMLQPEQLCMISGWGATEE 180
 Db 121 KYISHPNYDSKTKNNDIALMKLOKPLTFNDLVKVCPLPFGMLQPEQLCMISGWGATEE 180
 Oy 181 KGIETSEVINAARVLLIETORCNRSRYVDNLTTPAMICAGFLQGNVDSGQDSGGLVTSK 240
 Db 181 KGIETSEVINAARVLLIETORCNRSRYVDNLTTPAMICAGFLQGNVDSGQDSGGLVTSK 240
 Oy 241 NNIMWLIGDTSWGSACAKAYRPGYGNVMTFTDIYRQMBADG 283
 Db 241 NNIMWLIGDTSWGSACAKAYRPGYGNVMTFTDIYRQMBADG 283
 RESULT 3
 ID AAB14615 standard; Protein; 283 AA.
 XX AAB14615;
 AC
 XX

DT	18-JUN-2002	(first entry)
XX	Human prostate-associated protease (HUPAP).	
DE	Human; prostate-associated protease; HUPAP; serine protease; kallikrein;	
XX	gastrointestinal disorder; pancreatitis; ulcerative colitis; cancer;	
KW	adenocarcinoma; prostatic hyperplasia; prostate disorder.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Region	22..45
FT	/note= "Antigenic fragment. This region is	
FT	specifically claimed in claim 5 of the specification"	
FT	Active-site	87
FT	/note= "Essential for serine protease activity"	
FT	Active-site	136
FT	/note= "Essential for serine protease activity"	
FT	Active-site	232
FT	/note= "Essential for serine protease activity"	
FT	Misc-difference	235
FT	/label= "Unknown	
FT	/note= "Encoded by CMT"	
XX	US6350448-B1.	
XX		
PD	26-FEB-2002.	
XX		
PP	07-JAN-2000; 2000US-0478957.	
XX		
PR	27-FEB-1997; 97US-0807151.	
XX		
PA	(INCY-) INCYTE GENOMICS INC.	
XX		
PI	Bandman O, Lal P;	
DR	WPI: 2002-215263/27.	
DR	N-PSDB; AAD27930.	
XX		
PT	A purified human prostate-associated protease useful for treating a	
PT	gastrointestinal disorder, screening for a compound or molecule that	
PT	binds to it and for purifying such compounds -	
PS	Claim 1; Fig 1; 27pp; English.	
XX		
CC	The present sequence is human prostate-associated protease (HUPAP),	
CC	a serine protease belonging to kallikrein family. The HUPAP or its	
CC	agonist is used for treating gastrointestinal disorders (e.g.	
CC	pancreatitis, ulcerative colitis). The HUPAP antagonist can be used to	
CC	treat cancer (e.g. adenocarcinoma, colon cancer), and prostate disorders	
CC	e.g. prostatic hyperplasia. The HUPAP polypeptide can also be	
XX	used to screen a library of compounds that bind to it.	
XX		
ST	Sequence 283 AA;	
XX		
QY	Query Match 99.9%; Score 1541; DB 23; Length 283;	
Db	Best Local Similarity 100.0%; Pred. No. 3.7e-136;	
QY	Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
Db	1 MKLNTASGNVDIYKXLYHSDACSSKAVVSLRCLAGCVNLSSRSRLVYGESLPGMWP 60	
QY	1 MKLNTASGNVDIYKXLYHSDACSSKAVVSLRCLAGCVNLSSRSRLVYGESLPGMWP 60	
Db	1 MKLNTASGNVDIYKXLYHSDACSSKAVVSLRCLAGCVNLSSRSRLVYGESLPGMWP 60	
QY	61 QVSLHVNVAHCCGSIITPEWITVAHCVKEDLNNPWHMTAFAGILROSFWFYAGYOV 120	
Db	61 QVSLHVNVAHCCGSIITPEWITVAHCVKEDLNNPWHMTAFAGILROSFWFYAGYOV 120	
QY	121 KVIASHNPVDSKTNNDIALMKLOKPLTFNDLVKPCVCLNPGMLOPRLCWSGQATTE 180	
Db	121 KVIASHNPVDSKTNNDIALMKLOKPLTFNDLVKPCVCLNPGMLOPRLCWSGQATTE 180	
QY	181 KGIETSLVAAKVLLIETORCNISRYVDNLITPAMICAGTQGVNDSQGDSSGXLVTSK 240	
Db	181 KGIETSLVAAKVLLIETORCNISRYVDNLITPAMICAGTQGVNDSQGDSSGXLVTSK 240	

Db	181	KGTSBVAANAVALILFRQKCSRYYDNLITPAMICAGPLQGNVDS	CGDSGGLVTSK	240
Cy	241	NNIWTILGDTSGSCAKARPGYGVNMTTDMITRYMRADG		283
Db	241	NNIWTILGDTSGSCAKARPGYGVNMTTDMITRYMRADG		283
	RESULT 4			
ID	AAU69962	standard; Protein; 393 AA.		
XX	AAU69962;			
XX	30-JAN-2002 (first entry)			
XX	Human prostate cDNA encoded protein #88.			
DE	Human prostate cancer; cytosolic; immunostimulant; tumour; immunogen.			
XX	Human, prostate cancer; cytosolic; immunostimulant; tumour; immunogen.			
XX	Homo sapiens.			
OS	WO200173032-A2.			
XX	04-OCT-2001.			
PD	27-MAR-2001; 2001WO-US09919.			
XX	27-MAR-2000; 2000US-0536857.			
XX	09-MAY-2000; 2000US-0568100.			
PR	12-MAY-2000; 2000US-0570737.			
PR	13-JUN-2000; 2000US-0593793.			
PR	27-JUN-2000; 2000US-0605783.			
PR	10-AUG-2000; 2000US-0636215.			
PR	29-AUG-2000; 2000US-0651236.			
PR	05-SEP-2000; 2000US-0657279.			
PR	02-OCT-2000; 2000US-0679426.			
PR	10-OCT-2000; 2000US-0685166.			
PA	(CORI-) CORIXA CORP.			
XX	Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;			
P1	Panger GR, Retter MW, Stolk JA, Day CH, Vedyck TS, Carter D;			
P1	Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;			
XX	WPI; 2001-639232/73.			
XX	New human prostate-specific polypeptides and polynucleotides useful for			
XX	the diagnosis and treatment of cancer, especially prostate cancer -			
XX	Claim 2; Page 574-575; 579pp; English.			
CC	The invention relates to isolated prostate-specific			
CC	polynucleotides, polypeptides, fusion proteins of the polypeptides,			
CC	antibodies raised against the polypeptides (or antigenic epitopes			
CC	derived from them) and antigen-presenting cells expressing the			
CC	polypeptides. The antibodies are useful for detecting the presence of			
CC	cancer, especially prostate cancer. The polypeptides, polynucleotides and			
CC	the antigen-presenting cells are useful for stimulating and/or expanding			
CC	T cells specific for a tumour protein, and for inhibiting the development			
CC	of cancer especially prostate cancer. Compositions comprising the			
CC	polynucleotide and/or polypeptide are useful for stimulating an immune			
CC	response, and for treating cancer. The oligonucleotide is useful for			
CC	detecting cancer. The present sequence is a prostate specific			
CC	polypeptide of the invention.			
XX	Sequence 393 AA;			
XX	Query Match 99.8%; Score 1540; DB 22; Length 393;			
XX	Best Local Similarity 99.8%; Fred. No. 7e-136;			
XX	Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
XX	1 MCLANTSAGNVDIYKYLHSDACSSKAVASJRCIACGVNLSKSHSRIVGSESLPGAMPF			60

Db 111 MKLNTSAGNDIYKYLHSDACSKAVYSLRCIACGVNLNSRSRIVGSHALPGAMP 170
 QY 61 QVSLHVNVAHVCSSITTPBMTVTAHCVKELNPNMHTAFAGILROSFMYGAGYQVE 120
 Db 171 QVSLHVNVAHVCSSITTPBMTVTAHCVKELNPNMHTAFAGILROSFMYGAGYQVE 230
 QY 121 KVISHPNYSKTKNDIALMKLOKPLTFNDLVKPCLPNPGMLOPEQLCWSGMGATEE 180
 Db 231 KVISHPNYSKTKNDIALMKLOKPLTFNDLVKPCLPNPGMLOPEQLCWSGMGATEE 290
 QY 181 KGKTSVLANAKVLLIETORCNSRYVDNLITPAMI CAGPLQGVNDSQCGDGGXLYTSK 240
 Db 291 KGKTSVLANAKVLLIETORCNSRYVDNLITPAMI CAGPLQGVNDSQCGDGGXLYTSK 350
 QY 241 NNIMWLI GDTSMGSCAKAYRPGVGNVWVFTDWTYRMRADG 283
 Db 351 NNIMWLI GDTSMGSCAKAYRPGVGNVWVFTDWTYRMRADG 393

RESULT 5

AA01317
 ID AA01317 standard; Protein; 393 AA.

AC AA01317;

DT 04-OCT-2001 (first entry)

DE P1000C partial amino acid sequence 100-492.

XX Human; prostate cancer; prostate-specific; diagnosis; vaccine;

KW cytostatic; gene therapy; metastasis.

OS Homo sapiens.

XX WO200151633-A2.

XX 19-JUL-2001.

XX 16-JAN-2001; 2001MO-US01574.

XX 14-JAN-2000; 2000US-0483672.

XX (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;

PI Kalos MD, Fanger GR, Day CH, Retter MM, Stolk DA, Sheiky YAM;

XX Wang A, Mesgher MJ;

XX WPI; 2001-425873/45.

XX New polynucleotide encoding a prostate-specific protein, for

PT diagnosing, monitoring and treating prostate cancer in a patient and

PT for use in vaccines -

PS Claim 2, Page 540-541; 543pp; English.

XX The present invention describes polynucleotide sequences (I) which encode

CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,

CC and can be used in vaccine production and gene therapy. (II), (II),

CC antibodies to (II), fusion proteins comprising (II), and isolated

CC T cells prepared using (I) or (II) are used treat cancer in a patient.

CC (I) and the antibodies are also used in the detection of cancer in a

CC patient. The cancer that is diagnosed or treated is particularly

CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or

CC (I) can be used for monitoring the progression of cancer in a patient.

CC (I) and (II) can also be used to improve diagnostic and therapeutic

CC methods for prostate cancer. They can indicate the level of metastasis

CC as well as the prostate volume. AAH93357 to AAH93944 and AA01115 to

CC AA01318 represent polynucleotide and amino acid sequences used in the

CC exemplification of the present invention.

XX Sequence 393 AA;

Query Match 99.8%; Score 1540; DB 22; Length 393;
 Best Local Similarity 99.6%; Pred. No. 7e-136;
 Matches 282; Conservative 0; Mismatches 1; Indels 0;
 Gaps 0;
 QY 1 MKLNTSAGNDIYKYLHSDACSKAVYSLRCIACGVNLNSRSRIVGSHALPGAMP 60
 Db 111 MKLNTSAGNDIYKYLHSDACSKAVYSLRCIACGVNLNSRSRIVGSHALPGAMP 170
 QY 61 QVSLHVNVAHVCSSITTPBMTVTAHCVKELNPNMHTAFAGILROSFMYGAGYQVE 120
 Db 171 QVSLHVNVAHVCSSITTPBMTVTAHCVKELNPNMHTAFAGILROSFMYGAGYQVE 230
 QY 121 KVISHPNYSKTKNDIALMKLOKPLTFNDLVKPCLPNPGMLOPEQLCWSGMGATEE 180
 Db 231 KVISHPNYSKTKNDIALMKLOKPLTFNDLVKPCLPNPGMLOPEQLCWSGMGATEE 290
 QY 181 KGKTSVLANAKVLLIETORCNSRYVDNLITPAMI CAGPLQGVNDSQCGDGGXLYTSK 240
 Db 291 KGKTSVLANAKVLLIETORCNSRYVDNLITPAMI CAGPLQGVNDSQCGDGGXLYTSK 350
 QY 241 NNIMWLI GDTSMGSCAKAYRPGVGNVWVFTDWTYRMRADG 283
 Db 351 NNIMWLI GDTSMGSCAKAYRPGVGNVWVFTDWTYRMRADG 393

RESULT 6

AB95422
 ID AB95422 standard; Protein; 393 AA.

AC AB95422;

DT 19-JUL-2002 (first entry)

DE Human P1000C fragment SEQ ID NO 934.

XX Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;

KW gene therapy.

OS Homo sapiens.

XX US2002022248-A1.

XX 21-FEB-2002.

XX 12-JAN-2001; 2001US-0759143.

XX 25-FEB-1997; 97US-0806099.

XX 01-AUG-1997; 97US-0904804.

XX 09-FEB-1998; 98US-0020956.

XX 25-FEB-1998; 98US-0030607.

XX 14-JUL-1998; 98US-0115453.

XX 23-SEP-1998; 98US-0159812.

XX 15-JAN-1999; 99US-0232149.

XX 09-APR-1999; 99US-0288946.

XX 13-JUL-1999; 99US-0352616.

XX 12-NOV-1999; 99US-0439313.

XX 18-NOV-1999; 99US-0443686.

XX 27-MAR-2000; 2000US-0483672.

XX 09-MAY-2000; 2000US-0568100.

XX 12-MAY-2000; 2000US-0570737.

XX 13-JUN-2000; 2000US-0593793.

XX 27-JUN-2000; 2000US-0605783.

XX 10-AUG-2000; 2000US-0636215.

XX 29-AUG-2000; 2000US-0651236.

XX 06-SEP-2000; 2000US-0657279.

XX 02-OCT-2000; 2000US-0679426.

XX 10-OCT-2000; 2000US-0685166.

XX (XUJ/) XU J.

XX (DILL/) DILLON D C.

XX (MITC/) MITCHAM J L.

XX (HARL/) HARLOCKER S L.

PA (JIAN/) JIANG Y.
 PA (KALO/) KALOS M D.
 PA (FANG/) FANGER G R.
 PA (RETT/) RETTER M W.
 PA (STOL/) STOLK J A.
 PA (DAYC/) DAY C H.
 PA (VEDV/) VEDVICK T S.
 PA (CART/) CARTER D.
 PA (LISX/) LI S X.
 PA (WANG/) WANG A.
 PA (SKEI/) SKEIKY Y A W.
 PA (HEPL/) HEPLER W T.
 PA (HEND/) HENDERSON R A.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
 XX
 DR WPI; 2002-255649/30.
 XX
 PT New prostate-specific polynucleotides for diagnosing and treating
 PT diseases, in particular prostate cancer, and as markers for the
 PT progression of cancer -
 XX
 PS Claim 2; SEQ ID NO 934; 87pp; English.
 XX
 CC The present invention provides prostate-specific coding sequences and
 CC their encoded proteins. These can be used in the diagnosis and treatment
 CC of cancers, particularly prostate cancer. The present sequence is a
 CC protein described in the invention.
 CC
 XX
 SQ Sequence 393 AA;
 Query Match 99.8%; Score 1540; DB 23; Length 393;
 Best Local Similarity 99.6%; Pred. No. 7e-136; 1; Indels 0; Gaps 0;
 Matches 282; Conservative 0; Mismatches 1;
 QY 1 MKLNTSAGNVDIYKLYHSDACSSKAIVSLRCLACGVNLNSRGRIVGSSALPGAMPW 60
 DB 111 MKLNTSAGNVDIYKLYHSDACSSKAIVSLRCLACGVNLNSRGRIVGSSALPGAMPW 170
 QY 61 OVSILHNVNVAHVGSGSITTPFWITTAHCVCKPLNPNMHTAAGILRSGFMYGAGYVE 120
 DB 171 QVSLHVNVAHVGSGSITTPFWITTAHCVCKPLNPNMHTAAGILRSGFMYGAGYVE 230
 QY 121 KVISHPYVDSKTKQNDIALMKLOKPLTFNDLYVPVCLPYPGWMLQPEQLCWISGMGATEE 180
 DB 231 KVISHPYVDSKTKQNDIALMKLOKPLTFNDLYVPVCLPYPGWMLQPEQLCWISGMGATEE 290
 QY 181 KCKTSEVLNAKVLLETCRCNSRYRYDMLITPAMI CAGFLQGNVDSGQDSGGLVTSK 240
 DB 291 KCKTSEVLNAKVLLETCRCNSRYRYDMLITPAMI CAGFLQGNVDSGQDSGGLVTSK 350
 QY 241 NNIMWLIGDTSWGSGCAKARPGVYGNVWVFTDMITRORADG 283
 DB 351 NNIMWLIGDTSWGSGCAKARPGVYGNVWVFTDMITRORADG 393
 RESULT 7
 ABU71853 standard; Protein; 393 AA.
 XX
 AC ABU71853;
 XX
 DT 10-JUN-2003 (first entry)
 XX
 DE Prostate cancer associated protein #69.
 XX
 KW Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
 KW immunogen; cancer; prostate specific antigen; PSA;
 KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
 KW PSA.

OS Homo sapiens.
 XX
 XX US2002192763-A1.
 XX
 PD 19-DEC-2002.
 XX
 XX 29-JUN-2001; 2001US-0895793.
 XX
 XX 17-APR-2000; 2000US-157455P.
 XX 04-OCT-2000; 2000US-0679272.
 XX 28-MAR-2001; 2001US-0822827.
 XX
 XX (XUJ/) XU J.
 PA (DILL/) DILLON D C.
 PA (MITC/) MITCHAM J L.
 PA (HARL/) HARLOCKER S L.
 PA (JIAN/) JIANG Y.
 PA (KALO/) KALOS M D.
 PA (FANG/) FANGER G R.
 PA (RETT/) RETTER M W.
 PA (STOL/) STOLK J A.
 PA (DAYC/) DAY C H.
 PA (VEDV/) VEDVICK T S.
 PA (CART/) CARTER D.
 PA (LISX/) LI S X.
 PA (WANG/) WANG A.
 PA (SKEI/) SKEIKY Y A W.
 PA (HEPL/) HEPLER W T.
 PA (HEND/) HENDERSON R A.
 PA (HURA/) HURAL J.
 PA (MCNE/) MCNEILL P D.
 PA (HONG/) HUGHTON R L.
 PA (DEAS/) Y DE BASSOLS C V.
 PA (FOYT/) FOY T M.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA, Hural J;
 PI Mcneill PD, Houghton RL, Y De Bassols CV, Foy TM;
 XX
 DR WPI; 2003-352711/33.
 XX
 PT New fusion protein comprising prostate-specific polypeptides, or its
 PT immunogenic portions, useful for diagnosing, preventing and/or treating
 PT cancer, particularly prostate cancer -
 XX
 PS Example 2; SEQ ID NO 934; 85pp; English.
 XX
 CC The invention describes a fusion protein comprising at least one amino
 CC acid sequence of immunogenic portions of any of the 3 sequences not
 CC defined in the specification, or sequences having at least 70 or 90 %
 CC sequence identity to any one of the 35 sequences defined in the USPTO
 CC web site, which is encoded by any of the 4 nucleotide sequences not
 CC defined in the specification. The fusion protein, composition and
 CC methods are useful for diagnosing, preventing and/or treating cancer,
 CC particularly prostate cancer. The proteins are useful as markers to
 CC indicate the presence or absence of cancer. This is the amino acid
 CC sequence of a prostate cancer therapy associated protein.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from the US patent office at
 CC seqdata.uspto.gov/sequence.html?DocID=US20020192763.
 CC
 XX
 SQ Sequence 393 AA;
 Query Match 99.8%; Score 1540; DB 24; Length 393;
 Best Local Similarity 99.6%; Pred. No. 7e-136; 1; Indels 0; Gaps 0;
 Matches 282; Conservative 0; Mismatches 1;
 QY 1 MKLNTSAGNVDIYKLYHSDACSSKAIVSLRCLACGVNLNSRGRIVGSSALPGAMPW 60
 DB 111 MKLNTSAGNVDIYKLYHSDACSSKAIVSLRCLACGVNLNSRGRIVGSSALPGAMPW 170

QY 61 QVSLHVNQVHVCSSITTEPMITVTAHCEKELNPNPMTAFAGILRQSEMFYAGYQVE 120
Db 171 QVSLHVNQVHVCSSITTEPMITVTAHCEKELNPNPMTAFAGILRQSEMFYAGYQVE 210
QY 121 KVISHPYDSKTNNDIALMCKQKPLTENDLVKPYCLPNPGMTLQPBOLCWSGMAATEE 180
Db 231 KVISHPYDSKTNNDIALMCKQKPLTENDLVKPYCLPNPGMTLQPBOLCWSGMAATEE 230
QY 181 KGTSEVTLNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNVDSQCGDSGGLVTSK 240
Db 291 KGTSEVTLNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNVDSQCGDSGGLVTSK 350
QY 241 NNIMWLIGDTSWGSQCAKAYRPGVGNVWVFTDWTYRQMRADG 283
Db 351 NNIMWLIGDTSWGSQCAKAYRPGVGNVWVFTDWTYRQMRADG 393

RESULT 8
AAV92050
ID AAV92050 standard; Protein; 492 AA.
XX
AC AAV92050;
XX
DT 01-AUG-2000 (first entry)
XX
DE H-PCa6/7 polypeptide from androgen-inducible gene clone.
XX
KW Androgen inducible; testosterone; prostate cancer; cytostatic;
XX
OS TMPRSS2; diagnosis.
XX
OS Homo sapiens.
XX
FN WO200018961-A2.
XX
PD 06-APR-2000.
XX
PF 30-SEP-1999; 99MO-US22535.
XX
PR 30-SEP-1998; 98US-0163759.
XX
PR 30-SEP-1998; 98US-0164159.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Macbeth KJ, Shyjan AW;
XX
DR WPI, 2000-293182/25.
XX
DR N-PSDB; AAA08803.
XX
PT Novel methods for identifying compounds for treating prostate cancer
XX
PS comprising measuring the level of expression or activity of 1 or more
XX
PS of 11 genes or their products
XX
PS Claim 2; Fig 3; 108pp; English.
XX
CC This protein is encoded by a gene which is androgen (e.g. testosterone)
XX
CC inducible in androgen-dependent prostate cancer cells (e.g. LNCaP cells)
XX
CC and constitutively expressed in androgen-independent prostate cancer
XX
CC cells (e.g. LNCaP cells). Agents which decrease the expression or
XX
CC activity of these clones may slow or arrest the growth of prostate cancer
XX
CC cells or may kill them. H-PCa6/7 can be obtained from the sequence of
XX
CC the known gene for TMPRSS2. A compound useful for treating prostate
XX
CC cancer can be identified in a novel method comprising measuring the
XX
CC expression level, or activity, of H-PCa2, 3, 6/7, 8, 9, 10, 13, 14, 15,
XX
CC 19, or peripherol-type benzodiazepine receptor (PBR) in a cell, in the
XX
CC presence and absence of a test compound. The sequences may also be used
XX
CC in diagnosis of prostate cancer and to determine efficacy of treatment
XX
CC for prostate cancer.
XX
SQ Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 21; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKNTSAGNVYDIKCLYHSDACSSRAVYSLRCLACGVNLSSRQSRIVGSESLPGAMPW 60
Db 210 MKNTSAGNVYDIKCLYHSDACSSRAVYSLRCLACGVNLSSRQSRIVGSESLPGAMPW 269
QY 61 QVSLHVNQVHVCSSITTEPMITVTAHCEKELNPNPMTAFAGILRQSEMFYAGYQVE 120
Db 270 QVSLHVNQVHVCSSITTEPMITVTAHCEKELNPNPMTAFAGILRQSEMFYAGYQVE 329
QY 121 KVISHPYDSKTNNDIALMCKQKPLTENDLVKPYCLPNPGMTLQPBOLCWSGMAATEE 180
Db 330 KVISHPYDSKTNNDIALMCKQKPLTENDLVKPYCLPNPGMTLQPBOLCWSGMAATEE 389
QY 181 KGTSEVTLNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNVDSQCGDSGGLVTSK 240
Db 390 KGTSEVTLNAKVLLETQRCNSRYVDNLITPAMICAGFLQGNVDSQCGDSGGLVTSK 449
QY 241 NNIMWLIGDTSWGSQCAKAYRPGVGNVWVFTDWTYRQMRADG 283
Db 450 NNIMWLIGDTSWGSQCAKAYRPGVGNVWVFTDWTYRQMRADG 492

RESULT 9
AAV77726
ID AAV77726 standard; Protein; 492 AA.
XX
AC AAV77726;
XX
DT 12-MAY-2000 (first entry)
XX
DE Human tumour suppressor TMPRSS2 polypeptide.
XX
KW Tumour suppressor gene; TMPRSS2; cancer; human; drug design;
XX
KW gene therapy; protein therapy.
XX
OS Homo sapiens.
XX
FN WO200000605-A1.
XX
PD 06-JAN-2000.
XX
PF 29-JUN-1999; 99MO-US14622.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PR 29-JUN-1998; 98US-0091044.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Wong AKC, Tavtigian SV, Teng DHF;
XX
DR WPI, 2000-170914/15.
XX
DR N-PSDB; AAZ87786.
XX
PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
XX
PT human cancer -
XX
PS Claim 55; Page 77-79; 89pp; English.
XX
CC The invention provides a new tumour suppressor gene, designated TMPRSS2.
XX
CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for
XX
CC diagnosing and prognosing predisposition to cancer in humans. The
XX
CC polypeptides may also be used in assays to screen for compounds with
XX
CC anti-cancer or therapeutic properties. The polypeptides are also useful
XX
CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides
XX
CC may be used for gene therapy and protein therapy. The present sequence
XX
CC represents the TMPRSS2 polypeptide.
XX
SQ Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 21; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKNTSAGNVYDIKCLYHSDACSSRAVYSLRCLACGVNLSSRQSRIVGSESLPGAMPW 60

Db 210 MKLNTSAGNDIYKQIYHSDAGSKAVVSLRCLAGVNLNSRGRIVGSGSALPGAMPW 269
 Qy 61 QVSLHVNQVAVCGGSIITPEMIVTAHCEKELNPMWMTAFAGILROSFMFYAGYQVE 120
 Db 270 QVSLHVNQVAVCGGSIITPEMIVTAHCEKELNPMWMTAFAGILROSFMFYAGYQVE 329
 Qy 121 KVISHPYDSKTNNDIALMKLOKPLETENDLVKPCLPNPGMQLQPEOLCMIISGMGATEE 180
 Db 330 KVISHPYDSKTNNDIALMKLOKPLETENDLVKPCLPNPGMQLQPEOLCMIISGMGATEE 389
 Qy 181 KGTSEVLANAKVLLIETORCNSRYVDNLITPAMICAGFLQGNVDSQGDGSGGLVTSK 240
 Db 390 KGTSEVLANAKVLLIETORCNSRYVDNLITPAMICAGFLQGNVDSQGDGSGGLVTSK 449
 Qy 241 NNTIWMILIGDTSWGSCKAKAYRPGVGNVWFETDWTYRQMRADG 283
 Db 450 NNTIWMILIGDTSWGSCKAKAYRPGVGNVWFETDWTYRQMRADG 492

RESULT 10

AAV44406 standard; Protein; 492 AA.

AAV44406;

22-MAR-2000 (first entry)

Human 20P1F12-GTC2 protein.

20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer; transmembrane protein; colon; prostate; prostate tumour.

Homo sapiens.

MO9962942-A2.

09-DEC-1999.

01-JUN-1999; 99MO-US12253.

01-JUN-1998; 98US-0087598.

29-JUN-1998; 98US-0091474.

14-APR-1999; 99US-0129521.

(UROG-) UROGENESYS INC.

(AFAR/) AFAR D E.

(HUBE/) HUBERT R S.

(LEON/) LEONG K.

(RAIT/) RAITANO A B.

(SAFE/) SAFRAN D C.

Afar DE, Hubert RS, Leong K, Raitano AB, Safran DC;

WPI; 2000-116363/10.

N-PSDB; AAZ29636.

Novel cell surface antigen useful to treat colon and prostate cancer -

Claim 1; Fig 1; 58pp; English.

The present sequence is the 20P1F12 protein (also known as the TMPRSS2 protein) which is a prostate-specific, androgen-regulated, cell surface serine protease. It is a glycosylated type II transmembrane protein with an extracellular C-terminal serine protease domain, a scavenger receptor cysteine-rich domain, an LDL receptor class A domain and a predicted transmembrane domain. Host cells can be transformed to produce this protein, using vector containing 20P1F12/TMPRSS2 gene (also designated 20P1F12-GTC1, as deposited with ATCC accession number 2070977). Anti-20P1F12/TMPRSS2 antibodies may be used as therapeutic agent for prostate and colon cancers, to image prostate cancer cells and prostate tumours, to identify ligands and cellular constituents that bind to a 20P1F12/TMPRSS2 gene product and for use as cancer vaccines.

XX Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 21; Length 492;
 Best Local Similarity 99.6%; Pred. No. 9.2e-136;
 Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLNTSAGNDIYKQIYHSDAGSKAVVSLRCLAGVNLNSRGRIVGSGSALPGAMPW 60
 Db 210 MKLNTSAGNDIYKQIYHSDAGSKAVVSLRCLAGVNLNSRGRIVGSGSALPGAMPW 269
 Qy 61 QVSLHVNQVAVCGGSIITPEMIVTAHCEKELNPMWMTAFAGILROSFMFYAGYQVE 120
 Db 270 QVSLHVNQVAVCGGSIITPEMIVTAHCEKELNPMWMTAFAGILROSFMFYAGYQVE 329
 Qy 121 KVISHPYDSKTNNDIALMKLOKPLETENDLVKPCLPNPGMQLQPEOLCMIISGMGATEE 180
 Db 330 KVISHPYDSKTNNDIALMKLOKPLETENDLVKPCLPNPGMQLQPEOLCMIISGMGATEE 389
 Qy 181 KGTSEVLANAKVLLIETORCNSRYVDNLITPAMICAGFLQGNVDSQGDGSGGLVTSK 240
 Db 390 KGTSEVLANAKVLLIETORCNSRYVDNLITPAMICAGFLQGNVDSQGDGSGGLVTSK 449
 Qy 241 NNTIWMILIGDTSWGSCKAKAYRPGVGNVWFETDWTYRQMRADG 283
 Db 450 NNTIWMILIGDTSWGSCKAKAYRPGVGNVWFETDWTYRQMRADG 492

RESULT 11

AAU69960 standard; Protein; 492 AA.

AAU69960;

30-JAN-2002 (first entry)

Human prostate cDNA encoded protein #86.

Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.

Homo sapiens.

MO200173032-A2.

04-OCT-2001.

27-MAR-2001; 2001MO-US09919.

27-MAR-2000; 2000US-0536857.

09-MAY-2000; 2000US-0568100.

12-MAY-2000; 2000US-0570737.

13-JUN-2000; 2000US-0593793.

27-JUN-2000; 2000US-0605783.

10-AUG-2000; 2000US-0636215.

29-AUG-2000; 2000US-0651236.

06-SEP-2000; 2000US-0657279.

02-OCT-2000; 2000US-0679426.

10-OCT-2000; 2000US-0685166.

(CORI-) CORIXA CORP.

Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kates MD, Ranger GR, Ketter MW, Stolk JA, Day CH, Vedvik TS, Carter D; Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;

WPI; 2001-639232/73.

N-PSDB; AAS64178, AAS64179, AAS64180.

New human prostate-specific polypeptides and polynucleotides useful for the diagnosis and treatment of cancer, especially prostate cancer -

Claim 2; Page 573-574; 579pp; English.

CC The invention relates to isolated prostate-specific
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC antibodies raised against the polypeptides (or antigenic epitopes
CC derived from them) and antigen-presenting cells expressing the
CC polypeptides. The antibodies are useful for detecting the presence of
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC the antigen-presenting cells are useful for stimulating and/or expanding
CC T cells specific for a tumor protein, and for inhibiting the development
CC of cancer especially prostate cancer. Compositions comprising the
CC polynucleotide and/or polypeptide are useful for stimulating an immune
CC response, and for treating cancer. The oligonucleotide is useful for
CC detecting cancer. The present sequence is a prostate specific
CC polypeptide of the invention.

CC
XX
SQ Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 22; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIAGVNLSSRSRIVGESALPGAMPW 60
DB 210 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIAGVNLSSRSRIVGESALPGAMPW 269
QY 61 QVSLHQNHHVCGGSIITPEWIVTAHCEKPLNNPMTAFAGILFQSMFYGAGYQVE 120
DB 270 QVSLHQNHHVCGGSIITPEWIVTAHCEKPLNNPMTAFAGILFQSMFYGAGYQVE 329
QY 121 KVSHPNYSKTKNDIALMKLQKPLTFNDLVKPCJPNPMTLQPBOLCMISGMGATEE 180
DB 330 KVSHPNYSKTKNDIALMKLQKPLTFNDLVKPCJPNPMTLQPBOLCMISGMGATEE 389
QY 181 KGTSTVLAARVLLIETRCNSRYVDNLITPAMICAGFLQGVNDSGQDSGGLVTSK 240
DB 390 KGTSTVLAARVLLIETRCNSRYVDNLITPAMICAGFLQGVNDSGQDSGGLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAKYRPGVYGNVWFTDVIYRQMRADG 283
DB 450 NNIMWLIGDTSWGSCKAKYRPGVYGNVWFTDVIYRQMRADG 492

RESULT 12

AA001315 standard; Protein; 492 AA.

AA001315;

04-OCT-2001 (first entry)

P1000C amino acid sequence.

Human; prostate cancer; prostate-specific; diagnosis; vaccine;

cytostatic; gene therapy; metastasis.

Homo sapiens.

WO200151633-A2.

19-JUL-2001.

16-JAN-2001; 2001WO-US01574.

14-JAN-2000; 2000US-0483672.

(CORI-) CORIXA CORP.

Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG,

Kaloupek MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skelky YAW,

WPI; 2001-425873/45.

New polynucleotide encoding a prostate-specific protein, for

PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -

XX Claim 2; Page 539-540; 543pp; English.

XX The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used to treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnosis and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AA093357 to AA093944 and AA001115 to
CC AA001318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.

CC
XX
SQ Sequence 492 AA;

Query Match 99.8%; Score 1540; DB 22; Length 492;
Best Local Similarity 99.6%; Pred. No. 9.2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIAGVNLSSRSRIVGESALPGAMPW 60
DB 210 MKLNTSAGNVDIYKLYHSDACSSKAVVSLRCIAGVNLSSRSRIVGESALPGAMPW 269
QY 61 QVSLHQNHHVCGGSIITPEWIVTAHCEKPLNNPMTAFAGILFQSMFYGAGYQVE 120
DB 270 QVSLHQNHHVCGGSIITPEWIVTAHCEKPLNNPMTAFAGILFQSMFYGAGYQVE 329
QY 121 KVSHPNYSKTKNDIALMKLQKPLTFNDLVKPCJPNPMTLQPBOLCMISGMGATEE 180
DB 330 KVSHPNYSKTKNDIALMKLQKPLTFNDLVKPCJPNPMTLQPBOLCMISGMGATEE 389
QY 181 KGTSTVLAARVLLIETRCNSRYVDNLITPAMICAGFLQGVNDSGQDSGGLVTSK 240
DB 390 KGTSTVLAARVLLIETRCNSRYVDNLITPAMICAGFLQGVNDSGQDSGGLVTSK 449
QY 241 NNIMWLIGDTSWGSCKAKYRPGVYGNVWFTDVIYRQMRADG 283
DB 450 NNIMWLIGDTSWGSCKAKYRPGVYGNVWFTDVIYRQMRADG 492

RESULT 13

ABB95420 standard; Protein; 492 AA.

ABB95420;

19-JUL-2002 (first entry)

Human P1000C SEQ ID NO 932.

Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;

gene therapy.

Homo sapiens.

US2002022248-A1.

21-FEB-2002.

12-JAN-2001; 2001US-0759143.

25-FEB-1997; 97US-0806099.

01-AUG-1997; 97US-0904804.

09-FEB-1998; 98US-0020956.

25-FEB-1998; 98US-0030607.

14-JUL-1998; 98US-0115453.


```

PR 23-SEP-1998; 98US-0159812.
PR 15-JAN-1999; 99US-0232149.
PR 09-APR-1999; 99US-0288946.
PR 13-JUL-1999; 99US-0352616.
PR 12-NOV-1999; 99US-0439313.
PR 18-NOV-1999; 99US-0443686.
PR 14-JAN-2000; 2000US-0483672.
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
XX (XUJ)/ XU J.
PA (DILL)/ DILLON D C.
PA (MITC)/ MITCHEAM J L.
PA (HARL)/ HARLOCKER S L.
PA (JIAN)/ JIANG Y.
PA (KALO)/ KALOS M D.
PA (FANG)/ FANGER G R.
PA (RETT)/ RETTER M W.
PA (STOL)/ STOLK J A.
PA (DAYC)/ DAY C H.
PA (VEDV)/ VEDVICK T S.
PA (CART)/ CARTER D.
PA (LISX)/ LI S X.
PA (WANG)/ WANG A.
PA (SKEI)/ SKEIKY Y A W.
PA (HEPL)/ HEPLER W T.
PA (HEND)/ HENDERSON R A.
XX
XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD,
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D,
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA,
XX WPI; 2002-255649/30.
XX
XX New prostate-specific polynucleotides for diagnosing and treating
PT diseases, in particular prostate cancer, and as markers for the
PT progression of cancer.
XX
XX Claim 2; SEQ ID NO 932; 87bp; English.
XX
XX The present invention provides prostate-specific coding sequences and
CC their encoded proteins. These can be used in the diagnosis and treatment
CC of cancers, particularly prostate cancer. The present sequence is a
CC protein described in the invention.
XX
XX Sequence 492 AA;
SQ
Query Match 99.8%; Score 1540; DB 23; Length 492;
Best Local Similarity 99.6%; Pred. No. 9,2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MKLNTSAGNDVIYKJYHSDACSSKAVSLRCLAGVNLNSRGRIVGSSALPGAMPW 60
DB 210 MKLNTSAGNDVIYKJYHSDACSSKAVSLRCLAGVNLNSRGRIVGSSALPGAMPW 269
QY 61 QVSLHONVHVCSSITTEBWTATAHCEKPLNPMWTAFAAGILRGSFMYGAGYOVE 120
DB 270 QVSLHONVHVCSSITTEBWTATAHCEKPLNPMWTAFAAGILRGSFMYGAGYOVE 329
QY 121 KVISHPNVDKTYNDIALMKLOKPLTFNDLVKPCLPBPGMLOPBOLCWISGMGATEE 180
DB 330 KVISHPNVDKTYNDIALMKLOKPLTFNDLVKPCLPBPGMLOPBOLCWISGMGATEE 389
QY 181 KGITSEVLAARVLLLEFORCNSRYVDLITPAMICAGFLOGVNDSCOGDSGGKXATYTSK 240

```

```

DB 390 KGITSEVLAARVLLLEFORCNSRYVDLITPAMICAGFLOGVNDSCOGDSGGKXATYTSK 449
QY 241 NNITWMLIGDTSWGGCAGKAYRPGVYGNMVFETDVIYRQMRADG 283
DB 450 NNITWMLIGDTSWGGCAGKAYRPGVYGNMVFETDVIYRQMRADG 492
RESULT 14
AAE18096
ID AAE18096 standard; Protein; 492 AA.
XX
XX AAE18096;
AC
XX 07-MAY-2002 (first entry)
DT
XX Human 20P1F12-GTCL protein.
DB
XX
XX Serine protease; 20P1F12/TMPRSS2; 20P1F12-GTCL; cell growth; neoplasm;
XX cancer; vaccine; human.
XX
XX Homo sapiens.
OS
XX
XX MO200204953-A2.
PN
XX
XX 17-JAN-2002.
PD
XX
XX 12-JUL-2001; 2001WO-US22168.
PF
XX
XX 12-JUL-2000; 2000US-0615285.
PR
XX
XX (AGEN-) AGENSYS INC.
PA
XX
XX Saferran D, Raitano AB, Hubert RS, Jakobovits A, Paris M;
PI Chalfite-Bid PM;
PI N-PSDB; AAD28778.
XX
XX WPI; 2002-154967/20.
XX
XX Examining a biological sample for evidence of dysregulated cellular
PT growth, comprises comparing the status of prostate-specific,
PT androgen-regulated, secreted serine protease, 20P1F12/TMPRSS2, in a
PT corresponding normal sample.
XX
XX Claim 40; Fig 1; 161bp; English.
XX
XX The present invention relates to methods and compositions for the
CC diagnosis and therapy of prostate, colon, bladder, lung, ovarian and
CC kidney cancer derived from or based on a normally prostate-specific,
CC androgen regulated, cell membrane associated secreted serine protease
CC termed 20P1F12/TMPRSS2. The invention further relates to a method of
CC examining a biological sample for evidence of dysregulated cellular
CC growth comprising comparing the status of 20P1F12/TMPRSS2 gene (also
CC designated 20P1F12-GTCL) in the sample to the status of 20P1F12/TMPRSS2
CC in a corresponding normal sample. The invention also relates to 20P1F12/
CC TMPRSS2 polynucleotides and their corresponding proteins. Methods of the
CC invention are used for examining a sample such as blood, serum, stool,
CC urine, semen, or biopsy tissue for evidence of dysregulated cell growth.
CC The dysregulated cell growth is indicative of bladder cancer, lung
CC cancer, kidney cancer or ovarian cancer. It is useful for identifying
CC evidence of a neoplasm in a sample. Vaccines comprising an immunogenic
CC portion of 20P1F12/TMPRSS2 are useful for inhibiting growth of a cell
CC expressing 20P1F12/TMPRSS2 in a patient suffering from bladder cancer,
CC lung cancer, ovarian cancer or metastatic cancer. The present sequence
CC is human 20P1F12-GTCL protein.
XX
XX Sequence 492 AA;
SQ
Query Match 99.8%; Score 1540; DB 23; Length 492;
Best Local Similarity 99.6%; Pred. No. 9,2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MKLNTSAGNDVIYKJYHSDACSSKAVSLRCLAGVNLNSRGRIVGSSALPGAMPW 60

```

```

Db      210 MKLNTSAGNVDIYKQLYHSDACSSKAVVSLRCLACGVNLNSRSRIVGSSALPGAMPW 269
Qy      61 QVSLHVNQNVHVCSSGSIITPEMIVTAHCVKEKPLNPNMHTAFAGILRSFMYGAGYQVE 120
Db      270 QVSLHVNQNVHVCSSGSIITPEMIVTAHCVKEKPLNPNMHTAFAGILRSFMYGAGYQVE 329
Qy      121 KVISHPYDSTKRNNDIALMKLOKPLTFNDLVKPVCLPFGMMLQPEQLCWSMGATEE 180
Db      330 KVISHPYDSTKRNNDIALMKLOKPLTFNDLVKPVCLPFGMMLQPEQLCWSMGATEE 389
Qy      181 KGTSEVLNAKAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
Db      390 KGTSEVLNAKAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 449
Qy      241 NNIMWLIGDTSWSSGCAKAYRPGVGNVWFPTDVIYRORADG 283
Db      450 NNIMWLIGDTSWSSGCAKAYRPGVGNVWFPTDVIYRORADG 492

RESULT 15
ABU71851
ID      ABU71851 standard; Protein; 492 AA.
AC      ABU71851;
DT      10-JUN-2003 (first entry)
DE      Prostate cancer associated protein #67.
XX      Prostate cancer vaccine; gene therapy; cytostatic; fusion protein;
XX      immunogen; cancer; prostate specific antigen; PSA;
XX      prostatic acid phosphatase; PAP; prostate specific membrane antigen;
XX      PMSA.
XX      Homo sapiens.
XX      OS
XX      US2002192763-A1.
XX      19-DEC-2002.
XX      29-JUN-2001; 2001US-0895793.
XX      17-APR-2000; 2000US-157455P.
XX      04-OCT-2000; 2000US-0679272.
XX      28-MAR-2001; 2001US-0822837.
XX      (XUJ/) XU J.
XX      (DILL/) DILLON D C.
XX      (MITC/) MITCHELL J L.
XX      (HARL/) HARLOCKER S L.
XX      (JIAN/) JIANG Y.
XX      (KALO/) KALOS M D.
XX      (FANG/) FANGER G R.
XX      (RETT/) RETTER M W.
XX      (STOL/) STOLK J A.
XX      (DAYC/) DAY C H.
XX      (VEDV/) VEDVICK T S.
XX      (CART/) CARTER D.
XX      (LISX/) LI S X.
XX      (WANG/) WANG A.
XX      (SKEI/) SKEIKY Y A W.
XX      (HEPL/) HEPLER W T.
XX      (HEND/) HENDERSON R A.
XX      (HURA/) HURAL J.
XX      (MCNE/) MCNETTIL P D.
XX      (HOUN/) HOUGHTON R L.
XX      (DBAS/) Y DE BASSOLS C V.
XX      (FOYT/) FOY T M.
XX      (XU J, Dillon DC, Mitcham JL, Harlocker St, Jiang Y, Kalos MD;
XX      Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D,
XX      Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hurai J,
XX      McNeill PD, Houghton RL, Y De Bassols CV, Foy TM;

```

```

XX      WPI, 2003-352711/33.
XX      New fusion protein comprising prostate-specific polypeptides, or its
XX      immunogenic portions, useful for diagnosing, preventing and/or treating
XX      cancer, particularly prostate cancer.
XX      Example 2; SEQ ID NO 932; 85pp; English.
XX      The invention describes a fusion protein comprising at least one amino
XX      acid sequence of immunogenic portions of any of the 3 sequences not
XX      defined in the specification, or sequences having at least 70 or 90 %
XX      sequence identity to any one of the 35 sequences defined in the USPTO
XX      web site, which is encoded by any of the 4 nucleotide sequences not
XX      defined in the specification. The fusion protein, composition and
XX      methods are useful for diagnosing, preventing and/or treating cancer,
XX      particularly prostate cancer. The proteins are useful as markers to
XX      indicate the presence or absence of cancer. This is the amino acid
XX      sequence of a prostate cancer therapy associated protein.
XX      Note: The sequence data for this patent did not form part of the
XX      printed specification, but was obtained in electronic format directly
XX      from the US patent office at
XX      seqdata.uspto.gov/sequence.html?DocID=US20020192763.
XX      SQ
XX      Sequence 492 AA;

```

```

Query Match 99.8%; Score 1540; DB 24; Length 492;
Best Local Similarity 99.8%; Pred. No. 9,2e-136;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 MKLNTSAGNVDIYKQLYHSDACSSKAVVSLRCLACGVNLNSRSRIVGSSALPGAMPW 60
Db      210 MKLNTSAGNVDIYKQLYHSDACSSKAVVSLRCLACGVNLNSRSRIVGSSALPGAMPW 269
Qy      61 QVSLHVNQNVHVCSSGSIITPEMIVTAHCVKEKPLNPNMHTAFAGILRSFMYGAGYQVE 120
Db      270 QVSLHVNQNVHVCSSGSIITPEMIVTAHCVKEKPLNPNMHTAFAGILRSFMYGAGYQVE 329
Qy      121 KVISHPYDSTKRNNDIALMKLOKPLTFNDLVKPVCLPFGMMLQPEQLCWSMGATEE 180
Db      330 KVISHPYDSTKRNNDIALMKLOKPLTFNDLVKPVCLPFGMMLQPEQLCWSMGATEE 389
Qy      181 KGTSEVLNAKAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 240
Db      390 KGTSEVLNAKAVLLIETORCNSRYVDNLITPAMI CAGFLQGNVDS CGDSGGXLYTSK 449
Qy      241 NNIMWLIGDTSWSSGCAKAYRPGVGNVWFPTDVIYRORADG 283
Db      450 NNIMWLIGDTSWSSGCAKAYRPGVGNVWFPTDVIYRORADG 492

```

```

Search completed: August 1, 2003, 18:11:40
Job time : 46 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: August 1, 2003, 18:09:03 / Search time 11 Seconds
(without alignments)
1209.869 Million cell updates/sec

Title: US-09-988-975a-1
Sequence: 1 MKLNTSGNVNDIYKLYHSD.....VXGNVMTFDIYRQWRADG 263

Scoring table: BLOSUM62
Gap 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1540	92.0	492	1	TMS2_HUMAN
2	1264.5	82.0	490	1	TMS2_MOUSE
3	686.5	45.0	453	1	TMS3_MOUSE
4	686.5	44.5	454	1	TMS3_HUMAN
5	592.5	38.4	457	1	TMS3_HUMAN
6	575.5	37.3	455	1	TMS5_MOUSE
7	558.5	36.2	437	1	TMS4_HUMAN
8	555.5	36.0	638	1	KAL_HUMAN
9	544.1	35.1	638	1	KAL_PAT
10	536.5	34.8	1035	1	ENTK_BOVIN
11	528.5	34.3	1019	1	ENTK_HUMAN
12	526.5	34.1	638	1	KAL_MOUSE
13	520.5	33.7	417	1	HERP_HUMAN
14	520	33.7	625	1	FALL_HUMAN
15	518.5	33.6	418	1	HATT_HUMAN
16	518.5	33.6	855	1	ST14_MOUSE
17	513.5	33.3	1669	1	ENTK_MOUSE
18	512.5	33.2	852	1	ST14_HUMAN
19	504.5	32.7	815	1	PLMN_MOUSE
20	503.5	32.6	416	1	HERP_PAT
21	502.5	32.6	1034	1	ENTK_PIG
22	501.5	32.5	436	1	HERP_MOUSE
23	492	31.9	421	1	ACRO_HUMAN
24	491.5	31.9	324	1	TEST_MOUSE
25	490	31.8	1042	1	CORI_HUMAN
26	486.5	31.5	422	1	DESI_HUMAN
27	481.5	31.2	415	1	ACRO_PIG
28	480.5	31.1	436	1	ACRO_MOUSE
29	465.5	30.2	431	1	ACRO_RABIT
30	462	29.9	290	1	MPN_HUMAN
31	461.5	29.9	437	1	ACRO_PAT
32	460.5	29.8	311	1	TRYG_MOUSE
33	459	29.7	1113	1	CORI_MOUSE

ALIGNMENTS

RESULT 1	ID	TMS2_HUMAN	STANDARD	PRT	492 AA.
AC	015393	Q9BXX1			
DT	15-JUL-1998	(Rel. 36, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Transmembrane protease, serine 2 precursor (BC 3.4.21.-).				
GN	TPRSS2 OR PRSS10.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=9746814; PubMed=9325052;				
RA	Paoloni-Giacchino A., Chen H., Peltch M.C., Rosaler C., Antonarakis S.E.;				
RT	"Cloning of the TPRS2 gene, which encodes a novel serine protease with transmembrane, IDRA, and SRCR domains and maps to 21q22.3.";				
RT	Genomics 44:309-320(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=21309069; PubMed=11414763;				
RA	Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;				
RT	"Mutation analyses of 268 candidate genes in human tumor cell lines.";				
RT	Genomics 74:352-364(2001).				
RN	[3]				
RP	SEQUENCE FROM N.A. AND MUTAGENESIS.				
RX	MEDLINE=21139112; PubMed=11243404;				
RA	Raitano A.B., Jakobovits A.;				
RT	"Catalytic cleavage of the androgen-regulated TPRS2 protease results in its secretion by prostate and prostate cancer epithelia.";				
RT	Cancer Res. 61:1686-1692(2001).				
RN	[4]				
RP	TISSUE SPECIFICITY.				
RX	MEDLINE=21104370; PubMed=11169526;				
RA	Vahtera M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;				
RT	"Expression of transmembrane serine protease TPRS2 in mouse and human tissues.";				
RT	J. Pathol. 193:134-140(2001).				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY CLEAVAGE AND SECRETED.				
CC	- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.				
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.				
CC	- SIMILARITY: Contains 1 SRCR domain.				
CC	- SIMILARITY: Contains 1 LDL-receptor class A domain.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements).				

CC or send an email to license@lsb-eb.ch).

DR EMBL: U75329; AAC51784.1; -
DR EMBL: AF123453; AAD37117.1; -
DR EMBL: AF270487; AAK29280.1; -
DR HSP: P00763; IDPO.
DR MEROPS: S01.247; -
DR Genew: HGNC:11876; TMERS52.
DR MIM: 602060; -
DR GO: GO:0005887; C:intrigol to plasma membrane; TAS.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Ser_receptor_Ty.
DR InterPro: IPR001190; Stc_receptor.
DR Pfam: Pf00089; trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDLa; 1.
DR SMART: SM00202; SR; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS01209; LDLFA_1; 1.
DR PROSITE: PSS0068; LDLFA_2; 1.
DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
DR PROSITE: PSS0287; SRCR_2; 1.
DR PROSITE: PSS0240; TRYPsin DOM; 1.
DR PROSITE: PSS00134; TRYPsin HS; 1.
DR PROSITE: PSS00135; TRYPsin SR; 1.
KW Hydroxylase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Polymorphism.
FT CHAIN 1 255 TRANSMEMBRANE PROTEASE, SERINE 2, NON-CATALYTIC CHAIN.
FT CHAIN 256 492 TRANSMEMBRANE PROTEASE, SERINE 2, CATALYTIC CHAIN.
FT DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 112 149 LDL-RECEPTOR CLASS A. SRR.
FT DOMAIN 150 242 SERINE PROTEASE.
FT DOMAIN 256 492 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 296 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 345 CHARGE RELAY SYSTEM.
FT ACT_SITE 441 441 CLEAVAGE (POTENTIAL).
FT SITE 255 256 BY SIMILARITY.
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 172 231 BY SIMILARITY.
FT DISULFID 185 241 BY SIMILARITY.
FT DISULFID 244 365 INTERCHAIN (BY SIMILARITY).
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 449 449 K -> N (IN DBSNP:1056602).
FT MUTAGEN 255 255 /Ptd=VAR 011692.
FT MUTAGEN 441 441 R->Q: LOSS OF CLEAVAGE.
FT CONFLICT 160 160 S->A: LOSS OF ACTIVITY.
FT CONFLICT 242 242 M -> V (IN REF. 3).
FT CONFLICT 329 329 I -> L (IN REF. 1).
FT CONFLICT 489 491 E -> Q (IN REF. 1).
FT CONFLICT 491 491 RND -> RKN (IN REF. 1).
SO SEQUENCE 492 AA; 53891 MW; CAB44FD174A90768 CRC64;

Query Match 99.8%; Score 1540; DB 1; Length 492;
Best Local Similarity 99.6%; Pred. No. 1,5e-13;
Matches 283; Conservative 0; Mismatches 1; Indels 0; Gaps 0

Oy 1 MKLNTSGNDVYIKKKYHSDAGSSKRVSLRCICAGVNLNSRGRIRYGGSSALPGAPMP 60
Db 210 MKLNTSGNDVYIKKKYHSDAGSSKRVSLRCICAGVNLNSRGRIRYGGSSALPGAPMP 265
Oy 61 QVSLHVNQVHVCGGSIITPEWITAAHCYKPLNPMWHTAFAGILROSFNFYAGYQVE 120

```

Db 270 QVSLHGVAVHVCQGSIIITPEIIVTAACHCEKPLANPWHVTAAGILRQSPFVYAGAGYVE 329
Qy 121 KYISHPNPDSKTRKNDNALMKLQKPLTFNDLVYVCLPARGMQLQPELOCMISGMGATEE 180
Db 330 KYISHPNPDSKTRKNDNALMKLQKPLTFNDLVYVCLPARGMQLQPELOCMISGMGATER 389
Qy 181 KCKTSEVINAARVLLIETQRCSNRYVYDNLITPAMICAGFLQGNVDSQCGSGGXLVTSK 240
Db 390 KCKTSEVINAARVLLIETQRCSNRYVYDNLITPAMICAGFLQGNVDSQCGSGGGLVTSK 449
Qy 241 NNIIMWLIGDTSMGSGCAKAYRPGYGVGVWVFTDMITRQRBAQG 283
Db 450 NNIIMWLIGDTSMGSGCAKAYRPGYGVGVWVFTDMITRQRBAQG 492

RESULT 2
TMS2_MOUSE
ID TMS2_MOUSE STANDARD: PRT; 490 AA.
AC 09J08; 09J04; 09Y82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 26-FEB-2003 (Rel. 41, Last annotation update)
Db transmembrane protease, serine 2 (BC 3.4.21.-) (Epitheliasin) (Plasmi
Db transmembrane protein X).
TMRSS2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX MGI_Fastid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=BALE/C;
RX MEDLINE=21104370; Pubmed=11169526;
RA Vearala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMRSS2 in mouse and
human tissues.";
RL J. Pathol. 193:134-140(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C;
RA Jacquinet B.J., Rao N.V., Rao G.N., Hoidal J.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ACTIVATED BY
CLEAVAGE AND SECRETED. (By similarity).
CC - TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
CC - SIMILARITY: BELONGS TO PRETISSIN FAMILY S1.
CC - SIMILARITY: Contains 1 LDI-receptor class A domain.
CC - SIMILARITY: Contains 1 SRCR domain.
CC - SIMILARITY: Contains 1 LDI-receptor class A domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to licenses@ebi.ac.uk).
CC -----
DR EMBL/AF199362.1 AAF97867.1; -
DR EMBL/AF243500.1 AAF64186.1; -
DR EMBL/AF113596.1 AAF21308.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.247; -
DR MGD: MGI:1354381; Tmpres2.
DR InterPro: IPR001134; Chymotrypsin.
DR InterPro: IPR002172; LDI_receptor_A.
DR InterPro: IPR001254; Ser_protease_Try.

```

```

DR InterPro: IPR001190: Srcr_receptor.
DR Pfam: PF00057: ldl_recept_a; 1.
DR Pfam: PF00530: SRCR; 1.
DR Pfam: PF00089: trypsin; 1.
DR PRINTS: PR00722: CHYMOTRYPSIN.
DR SMART: SM00192; LDla; 1.
DR SMART: SM00020; SR; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS01209; LDLa_2; 1.
DR PROSITE: PS00068; LDLa_2; 1.
DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
DR PROSITE: PS50287; SRCR_2; 1.
DR PROSITE: PS50240; TRYPsin DOM; 1.
DR PROSITE: PS00134; TRYPsin HIS; 1.
DR PROSITE: PS00135; TRYPsin SER; 1.
KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor; Zymogen.
FT CHAIN 1 253
FT CHAIN 254 490
FT CHAIN 490 490
FT DOMAIN 1 83
FT TRANSMEM 84 104
FT DOMAIN 105 490
FT DOMAIN 111 149
FT DOMAIN 150 242
FT DOMAIN 254 490
FT ACT_SITE 294 294
FT ACT_SITE 343 343
FT ACT_SITE 439 439
FT SITE 253 253
FT DISUFPID 76 125
FT DISUFPID 119 138
FT DISUFPID 132 147
FT DISUFPID 171 230
FT DISUFPID 184 240
FT DISUFPID 243 363
FT DISUFPID 279 295
FT DISUFPID 408 424
FT CARBOHYD 435 463
FT CARBOHYD 111 111
FT CARBOHYD 212 212
FT CARBOHYD 474 474
FT CONFLICT 122 122
FT CONFLICT 178 178
FT CONFLICT 320 320
FT CONFLICT 474 474
SQ SEQUENCE 490 AA; 53479 MW; 07D2B01EA4D8A1A9 CRC64;

Query Match 82.0%; Score 1264.5; DB 1; Length 490;
Best Local Similarity 82.3%; Pred.No.1.le-106;
Matches 232; Conservative 20; Mismatches 26; Indels 1; Gaps 1;

QY 1 MKLTAGAVNDVYKLYHSDACSKAVVSLRCLAGVNLSSRSRIYGESEALRGAWPW 60
DB 209 MKLVSSGNVDLYKLYHSDSCSRVVSRLCTEGVR-SYRKSRIVGGANAPGMPW 267
QY 61 QVSLHGVNVAVCGGSIITPEWITVTAHCEYKELNNPMTAFAGILRSFMYGAGYOV 120
DB 268 QVSLHGVNVAVCGGSIITPEWITVTAHCEYKELNNPMTAFAGILRSFMYGAGYOV 327
QY 121 KVTSHENYDSTKRNNDIALMRLQKPLTFNDLVKPCLENPMPAQPLQCLVSGMGATEE 180
DB 328 KVTSHENYDSTKRNNDIALMRLQKPLTFNDLVKPCLENPMPAQPLQCLVSGMGATEE 387
QY 181 KGTSTEVLAARVLLITQRCSRYVDLITPATICAGLTQGVNDSCQDPSGGXIVTSK 240
DB 388 KGTSTEVLAARVLLITQRCSRYVDLITPATICAGLTQGVNDSCQDPSGGXIVTSK 447
QY 241 NNITWMLIGDTSWGSCKAKAYRPGVYGNVWVFTDVIYQMRAD 282
DB 448 NNITWMLIGDTSWGSCKAKAYRPGVYGNVWVFTDVIYQMRAD 489

```

```

RESULT 3
ID TM33 MOUSE STANDARD; PRT; 453 AA.
AC Q8K1T0; OSVDS0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 3 [EC 3.4.21.-].
GN TMPSR33.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION IN ENAC
RP CLEARVAGE.
RX PubMed=12393794;
RA Guipont M., Vassanau G., Wattenhofer M., Shibuya K., Vazquez M.,
RA Dougherty L., Scamuffa N., Guida E., Okui M., Rosier C., Hancock M.,
RA Buchet K., Raymond A., Hummler E., Marzella P.L., Kudoh J.,
RA Shalitz N., Scott H.S., Antonarakis S.E., Rosaler B.C.;
RT "The transmembrane serine protease (TMPSR33) mutated in deafness
RT DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro."
RT Hum. Mol. Genet. 11:2829-2836(2002).
CC -1- FUNCTION: Probable protease. Seems to be capable of activating
CC ENaC.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum.
CC -1- TISSUE SPECIFICITY: Expressed in the spiral ganglion, the cells
CC supporting the organ of Corti and the stria vascularis.
CC -1- PTM: Undergoes autolytic activation.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: AJ429216; CA22137.1; -
DR EMBL: AJ300738; CAC83350.1; -
DR HSSP: P00761; IANL.
DR MGD: MGI:2155445; Tmpsr33.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR002172; LDL_receptor_A.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR001190; Srcr_receptor.
DR Pfam: PF00057; ldl_recept_a; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00192; LDla; 1.
DR SMART: SM00020; SR; 1.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS01209; LDLa_2; 1.
DR PROSITE: PS50068; LDLa_2; 1.
DR PROSITE: PS50287; SRCR_2; 1.
DR PROSITE: PS50240; TRYPsin DOM; 1.
DR PROSITE: PS00134; TRYPsin HIS; 1.
DR PROSITE: PS00135; TRYPsin SER; 1.
KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor; Zymogen;
KW Endoplasmic reticulum.
FT CHAIN 1 48
FT CHAIN 49 69
FT TRANSMEM 1 48
FT DOMAIN 49 69
FT DOMAIN 70 453
FT DOMAIN 72 108
FT DOMAIN 104 205
FT DOMAIN 217 448

```

```

FT ACT SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 304 304 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 400 400 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 216 217 CLEAVAGE (POTENTIAL).
FT DISULFID 73 85 BY SIMILARITY.
FT DISULFID 79 98 BY SIMILARITY.
FT DISULFID 92 107 BY SIMILARITY.
FT DISULFID 129 194 BY SIMILARITY.
FT DISULFID 142 204 BY SIMILARITY.
FT DISULFID 207 324 BY SIMILARITY.
FT DISULFID 242 258 BY SIMILARITY.
FT DISULFID 338 406 BY SIMILARITY.
FT DISULFID 369 385 BY SIMILARITY.
FT DISULFID 396 424 BY SIMILARITY.
FT CARBOHYD 221 221 N-LINKED (GLUCNAC...) (POTENTIAL).
FT CARBOHYD 117 117 L->H (IN REF. 1; CAC83350).
FT CONFLICT 246 246 V->I (IN REF. 1; CAC83350).
SQ SEQUENCE 453 AA; 49491 MW; 1ABCB10AF6B1E6F CMC64;

Query Match 45.0%; Score 694; DB 1; Length 453;
Best Local Similarity 49.6%; Pred. No. 3.4e-55;
Matches 135; Conservative 39; Mismatches 92; Indels 6; Gaps 4;

QY 12 IYKRLYHSDACSSKAVSLRCLACGVNLNSSRQSRIVGSGALPGAMPQVSLHYQNVAY 71
DB 184 LHSYVMREGCTSGHVTLKCSACSTRGYS--PRIVGNNSSLTQPMQVSLQFGYHL 241
-QY 72 CGGSITPEWITAHCHVKKPLNPMWHTAFAGILRQSFMYGA-GVQVKVISHNVD 130
DB 242 CGGSVTPMITYAHACV-YLHYRKSITVOYGLV--SLHDSFVPSHLVKKIYHKKYR 298
QY 131 KTKNDIALMKLOKPELTPNDLVKPCLPNGMLQPEOLCWSGNGATEEKGTSYLVA 190
DB 299 KRLGNDIALMKLSEPLTFDETQPLCLPNSSENPFDGKLCWTSNGATEGCDASPLVNH 358
QY 191 AVLLIETORCSRKYVDLITPAMICAGFLQGVNDSGGSGGLVTSKNNIIMLIGDT 250
DB 359 AAVPLISNKICHRDVGIGIISPSMCLGAYLGAGVDSGGSGGPLVCGERRIMKLVGAT 418
QY 251 SWSGGCAKAYRPGYGVNVPFTMTYRQMRAD 282
DB 419 SFGIGCAEVNKGQVYTRISPLDMIHQLERD 450

RESULT 4
TMS3_HUMAN STANDARD; PRT; 454 AA.
ID TMS3_HUMAN
AC P57727;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transmembrane protease, serine 3 (BC 3.4.21.-) (Serine protease
DE TADG-12) (Tumor-associated differentially-expressed gene-12 protein).
GN TMPRSS3 OR TADG12 OR ECHOS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOPFORMS A AND T).
RC TISSUE=Ovarian carcinoma.
RX MEDLINB-20531358; PubMed11068177;
RA Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N.,
RA Wang Y., Parmley T.H., O'Brien T.J.;
RT "Ovarian tumor cells express a novel multi-domain cell surface serine
RT protease."
RL Biochim. Biophys. Acta 1502:337-350(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOPFORMS A, B AND D), AND VARIANT ILE-53.
RX MEDLINB-20587879; PubMed11137999;
RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chirast R.,
RA Guipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
RA Younus F., Mehdi S.O., Radhakrishna U., Papasavvas M.P., Gehrig C.,

```

```

RA Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamir B.,
RA Antonarakis S.B.;
RT "Insertion of beta-satellite repeats identifies a transmembrane
RT protease causing both congenital and childhood onset autosomal
RT recessive deafness."
RL Nat. Genet. 27:59-63(2001).
RN [3]
RP SUBCELLULAR LOCATION, AND FUNCTION IN ENAC CLEAVAGE.
RX PubMed=12393794;
RA Guipponi M., Vassilau G., Wattenhofer M., Shibuya K., Vazquez M.,
RA Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M.,
RA Buchet K., Raymond A., Hummer E., Marzella P.L., Kudoh J.,
RA Shimizu N., Scott H.S., Antonarakis S.B., Rossier C.;
RT "The transmembrane serine protease (TMPRSS3) mutated in deafness
RT DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro."
RL Hum. Mol. Genet. 11:2829-2836(2002).
RN [4]
RP VARIANTS DFNB8/DFNB10 CYS-251 AND LEU-404.
RX MEDLINB-21354482; PubMed11462234;
RA Maemond S., Antonarakis S.B., Schwede T., Ghorbel A.M., Grati M.,
RA Papasavvas M.P., Drita M., Elgied-Bouilla A., Wattenhofer M.,
RA Rossier C., Scott H.S., Ayadi H., Guipponi M.;
RT "Novel missense mutations of TMPRSS3 in two consanguineous Tunisian
RT families with non-syndromic autosomal recessive deafness."
RL Hum. Mutat. 18:101-108(2001).
RN [5]
RP VARIANTS DFNB8/DFNB10 TRP-109; PHE-194 AND ARG-407, AND VARIANTS
RP ILE-53; SER-111 AND VAL-253.
RX MEDLINB-21317610; PubMed11424922;
RA Ben-Yosef T., Wattenhofer M., Riazuddin S., Ahmed Z.M., Scott H.S.,
RA Kudoh J., Shibuya K., Antonarakis S.B., Bonne-Tamir B.,
RA Radhakrishna U., Naz S., Ahmed Z., Riazuddin S., Pandya A.,
RA Nance W.B., Wilcox E.R., Friedman T.B., Morell R.J.;
RT "Novel mutations of TMPRSS3 in four DFNB8/810 families segregating
RT congenital autosomal recessive deafness."
RL J. Med. Genet. 38:396-400(2001).
RN [6]
RP VARIANTS DFNB8/DFNB10 GLY-103, AND VARIANTS ASN-173 AND THR-426.
RX MEDLINB-21504597; PubMed11907649;
RA Wattenhofer M., Di Iorio V., Rabionet R., Dougherty L., Pamparos A.,
RA Schwede T., Montserrat-Sentis B., Ardones L., Illades T.,
RA Pasquandissegile A., D'Amelio M., Alwan S., Rossier C., Dahl H.-H.M.,
RA Petersen M.B., Bastivall X., Gasparini P., Scott H.S.;
RA Antonarakis S.B.;
RT "Mutations in the TMPRSS3 gene are a rare cause of childhood
RT nonsyndromic deafness in Caucasian patients."
RL J. Mol. Med. 80:124-131(2002).
RN [7]
RP FUNCTION: Probable protease. Seems to be capable of activating
RP ENAC.
CC [1] SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum.
CC [2] ALTERNATIVE PRODUCTS:
CC [3] Event-Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=P57727-1; Sequence=Displayed;
CC Name=B; Synonyms=C;
CC IsoId=P57727-2; Sequence=VSP_005391;
CC Name=D;
CC IsoId=P57727-3; Sequence=VSP_005392;
CC Name=E; Synonyms=Truncated; TADG-12V;
CC IsoId=P57727-4; Sequence=VSP_005393; VSP_005394;
CC [4] TISSUE SPECIFICITY: Expressed in many tissues including fetal
CC cochlea. Isoform T is found at increased levels in some
CC carcinomas.
CC [5] PTM: Undergoes autophosphorylation.
CC [6] DISEASE: Defects in TMPRSS3 are a cause of childhood-onset
CC autosomal recessive neurosensory deafness 8 (DFNB8) [MIM:601072].
CC [7] DISEASE: Defects in TMPRSS3 are a cause of congenital autosomal
CC recessive neurosensory deafness 10 (DFNB10) [MIM:605316].
CC [8] SIMILARITY: BELONGS TO PERTIDASE FAMILY S1.
CC [9] SIMILARITY: Contains 1 SKC domain.
CC [10] SIMILARITY: Contains 1 LDU-receptor class A domain.
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL: AF01380; AAC37012.1; -
 DR EMBL: AB038157; BAB20077.1; -
 DR EMBL: AB038158; BAB20078.1; -
 DR EMBL: AB038159; BAB20079.1; -
 DR EMBL: AB038160; BAB20080.1; -
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.079; -
 DR Genem: HGNC:11877; TMRSS3.
 DR MIM: 605511; -
 DR MIM: 601072; -
 DR MIM: 605316; -
 DR GO: GO:0016021; C: integral to membrane; NAS.
 DR GO: GO:0006508; P: proteolysis and peptidolysis; NAS.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL receptor A.
 DR InterPro: IPR001254; Ser. protease Try.
 DR InterPro: IPR001190; Src receptor.
 DR Pfam: PF00057; ldl_recept_a; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00192; LDLa; 1.
 DR SMART: SM00202; SR; 1.
 DR PROSITE: PS01209; LDLA_2; 1.
 DR PROSITE: PS00668; LDLA_2; 1.
 DR PROSITE: PS00420; SRCR_3; FALSE_NEG.
 DR PROSITE: PS0287; SRCR_2; 1.
 DR PROSITE: PS0287; TRYPsin; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_HIS; 1.
 DR HydroLase: Serine protease; Transmembrane; Signal-anchor; Zymogen;
 DR EndoLase: Serine protease; Transmembrane; Signal-anchor; Zymogen;
 DR Disease mutation: Polymorphism.
 KW DOMAIN 1 48
 FT TRANSMEM 49 69
 FT DOMAIN 70 454
 FT DOMAIN 72 108
 FT DOMAIN 109 205
 FT DOMAIN 217 454
 FT ACT_SITE 257 257
 FT ACT_SITE 304 304
 FT ACT_SITE 401 401
 FT SITE 216 217
 FT DISULFID 73 85
 FT DISULFID 79 98
 FT DISULFID 92 107
 FT DISULFID 129 134
 FT DISULFID 142 142
 FT DISULFID 207 204
 FT DISULFID 242 258
 FT DISULFID 338 407
 FT DISULFID 370 386
 FT DISULFID 397 425
 FT CARBOHYD 221 221
 FT VARSPLIC 1 127
 FT VARSPLIC 318 454
 FT VARSPLIC 261 293

EMBL: AF01380; AAC37012.1; -
 EMBL: AB038157; BAB20077.1; -
 EMBL: AB038158; BAB20078.1; -
 EMBL: AB038159; BAB20079.1; -
 EMBL: AB038160; BAB20080.1; -
 HSSP: P00763; IDPO.
 MEROPS: S01.079; -
 Genem: HGNC:11877; TMRSS3.
 MIM: 605511; -
 MIM: 601072; -
 MIM: 605316; -
 GO: GO:0016021; C: integral to membrane; NAS.
 GO: GO:0006508; P: proteolysis and peptidolysis; NAS.
 InterPro: IPR001314; Chymotrypsin.
 InterPro: IPR002172; LDL receptor A.
 InterPro: IPR001254; Ser. protease Try.
 InterPro: IPR001190; Src receptor.
 Pfam: PF00057; ldl_recept_a; 1.
 Pfam: PF00089; trypsin; 1.
 PRINTS: PR00722; CHYMOTRYPSIN.
 SMART: SM00192; LDLa; 1.
 SMART: SM00202; SR; 1.
 PROSITE: PS01209; LDLA_2; 1.
 PROSITE: PS00668; LDLA_2; 1.
 PROSITE: PS00420; SRCR_3; FALSE_NEG.
 PROSITE: PS0287; SRCR_2; 1.
 PROSITE: PS0287; TRYPsin; 1.
 PROSITE: PS00134; TRYPsin_HIS; 1.
 PROSITE: PS00135; TRYPsin_HIS; 1.
 HydroLase: Serine protease; Transmembrane; Signal-anchor; Zymogen;
 EndoLase: Serine protease; Transmembrane; Signal-anchor; Zymogen;
 Disease mutation: Polymorphism.
 DOMAIN 1 48
 TRANSMEM 49 69
 DOMAIN 70 454
 DOMAIN 72 108
 DOMAIN 109 205
 DOMAIN 217 454
 ACT_SITE 257 257
 ACT_SITE 304 304
 ACT_SITE 401 401
 SITE 216 217
 DISULFID 73 85
 DISULFID 79 98
 DISULFID 92 107
 DISULFID 129 134
 DISULFID 142 142
 DISULFID 207 204
 DISULFID 242 258
 DISULFID 338 407
 DISULFID 370 386
 DISULFID 397 425
 CARBOHYD 221 221
 VARSPLIC 1 127
 VARSPLIC 318 454
 VARSPLIC 261 293

EMBL: AF01380; AAC37012.1; -
 EMBL: AB038157; BAB20077.1; -
 EMBL: AB038158; BAB20078.1; -
 EMBL: AB038159; BAB20079.1; -
 EMBL: AB038160; BAB20080.1; -
 HSSP: P00763; IDPO.
 MEROPS: S01.079; -
 Genem: HGNC:11877; TMRSS3.
 MIM: 605511; -
 MIM: 601072; -
 MIM: 605316; -
 GO: GO:0016021; C: integral to membrane; NAS.
 GO: GO:0006508; P: proteolysis and peptidolysis; NAS.
 InterPro: IPR001314; Chymotrypsin.
 InterPro: IPR002172; LDL receptor A.
 InterPro: IPR001254; Ser. protease Try.
 InterPro: IPR001190; Src receptor.
 Pfam: PF00057; ldl_recept_a; 1.
 Pfam: PF00089; trypsin; 1.
 PRINTS: PR00722; CHYMOTRYPSIN.
 SMART: SM00192; LDLa; 1.
 SMART: SM00202; SR; 1.
 PROSITE: PS01209; LDLA_2; 1.
 PROSITE: PS00668; LDLA_2; 1.
 PROSITE: PS00420; SRCR_3; FALSE_NEG.
 PROSITE: PS0287; SRCR_2; 1.
 PROSITE: PS0287; TRYPsin; 1.
 PROSITE: PS00134; TRYPsin_HIS; 1.
 PROSITE: PS00135; TRYPsin_HIS; 1.
 HydroLase: Serine protease; Transmembrane; Signal-anchor; Zymogen;
 EndoLase: Serine protease; Transmembrane; Signal-anchor; Zymogen;
 Disease mutation: Polymorphism.
 DOMAIN 1 48
 TRANSMEM 49 69
 DOMAIN 70 454
 DOMAIN 72 108
 DOMAIN 109 205
 DOMAIN 217 454
 ACT_SITE 257 257
 ACT_SITE 304 304
 ACT_SITE 401 401
 SITE 216 217
 DISULFID 73 85
 DISULFID 79 98
 DISULFID 92 107
 DISULFID 129 134
 DISULFID 142 142
 DISULFID 207 204
 DISULFID 242 258
 DISULFID 338 407
 DISULFID 370 386
 DISULFID 397 425
 CARBOHYD 221 221
 VARSPLIC 1 127
 VARSPLIC 318 454
 VARSPLIC 261 293

FT VARSPLIC 294 454
 FT VARSPLIC 53 53
 FT VARSPLIC 103 103
 FT VARSPLIC 109 109
 FT VARSPLIC 111 111
 FT VARSPLIC 173 173
 FT VARSPLIC 194 194
 FT VARSPLIC 251 251
 FT VARSPLIC 253 253
 FT VARSPLIC 404 404
 FT VARSPLIC 407 407
 FT VARSPLIC 426 426
 FT VARSPLIC 46 54

Query Match 44.5%; Score 686.5; DB 1; Length 454;
 Best Local Similarity 50.2%; Pred. No. 16-54;
 Matches 137; Conservative 37; Mismatches 92; Indels 7; Gaps 5;

12 IYKRLHSDACSSKAVSLRCAAGVNLSSQSRIVGGBALPQAMPQVSLHVQNVV 71
 184 LHSVYRECGASGHVVTQCTACGHRGYS--SRIVGANNLSLQPMQASLOFQGYHL 241
 72 CGSGITPEWIVTAAICVKEKPLNPNMHTAFAGILRQSFMEFGA-GVOVEKVISHPYDS 130
 242 CGSGVITPEWIIINACV-YDLTKRSWTIOGVV--SLDNPAPSHLVKIVHSKTP 298
 131 KTRNDIATLKLQKPTTNDLVKPVCLPENGMTLPQOLCWSIGGATEE-KGRTSEVLN 189
 299 KRLGNIDALMKLAGLPTFEMHIQVCLPNSBENPDKGVCTTSCMGATBDGADSPVLN 358
 190 AAKVLLIFORCNSR-VVYNNLTTPAMICAGFLQGVNDS CGDSGGLVTSKNNIWWLIGD 249
 359 HAAVPLISNKCINHRVYTGIIISPMCAGYITGVSDCGDSGGLVCOERRMLKLVA 418
 250 TSMGSCAKARVPPGVYGNMVFDTWYRQPAD 282
 419 TSGRIGCAEVNKGVTYVTSPLDIHOMERD 451

RESULT 5
 ID TMS5_HUMAN STANDARD; PRT; 457 AA.
 AC Q9H3S3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transmembrane protease, serine 5 (EC 3.4.21.-) (Spinesin).
 GN TMRSS3.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX PubMed=11741986;
 RA Yamaguchi N., Ohui A., Yamada T., Nakazato H., Mitsu S.;
 RT "Spinesin/TMRSS3, a novel transmembrane serine protease, cloned from
 human spinal cord.";


```

KW Glycoprotein, Alternative splicing.
FT DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT
FT DOMAIN 71 455 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 112 207 SRCR.
FT DOMAIN 218 455 SERINE PROTEASE.
FT ACT_SITE 258 455 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 308 408 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 405 405 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT SITE 217 218 CLEAVAGE (POTENTIAL).
FT DISULFID 135 196 BY SIMILARITY.
FT DISULFID 148 206 BY SIMILARITY.
FT DISULFID 209 328 BY SIMILARITY.
FT DISULFID 243 259 BY SIMILARITY.
FT DISULFID 374 390 BY SIMILARITY.
FT CARBOHYD 401 429 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 1 144 Missing (in isoform 2).
FT VARSPPLIC 1 10 /FTid=VSP 005395.
FT VARSPPLIC 1 182 /FTid=VSP 005396.
FT VARSPPLIC 183 192 /FTid=VSP 005397.
FT VARSPPLIC 325 325 /FTid=VSP 005398.
FT CONFLICT 455 AA; 49632 NM; 509317896899AA C9C64;
SQ SEQUENCE
Query Match 37.3%; Score 575.5; DB 1; Length 455;
Best Local Similarity 42.0%; Pred. No. 1.7e-44;
Matches 108; Conservative 39; Mismatches 105; Indels 5; Gaps 3;
- QY 22 CSSKAVASVSLTACVNLNLSRQSRIVGSGSALPGAMPQVSLHVNQVCGSIITPEW 81
DB 196 CPEGRVLSLKSCSEGAR---PLASRIVGCGAVASRWQMSVNLGSHTCGASVLAPEW 252
QY 82 IYVAACVEK-PLANPMHTAPRAGILROSFMFYGAGYOYEVYISHPNYDSKTKNNIDALM 140
DB 253 VYVAACVMSVFRSLRSSRVRAGVSHGAVQHOQHTWKEKIPPLYSAGNHDYVAL 312
QY 141 KLOKPLTNDLVKPVCLPMPGMMLQPEOLCNISGKATE-EKGKTSVYLNAAKYLITETQ 199
DB 313 QKRTPIPFSDTVDAVGLPAKEQYFPWGSCWISGKHDPBHTHSSDTLOQTPVPLSTH 372
QY 200 RCRSRVYVNLITPAMICAGFLQGVNDSCQGDGSGXLYTSKNNIMWLIIGDTSWGSGCAKA 259
DB 373 LCNSSCMYSGALTHRMLCAGYLDGRADACQGDGSGPVLVCGSDTWHLVGVVSGRCAP 432
QY 260 YRPGVYGVNWFDTWY 276
DB 433 NRGVYAKVAEFLDWH 449

```

```

RC TISSUE=Pancreatic carcinoma;
RX MEDLINE=20283276; PubMed=10825129;
RA Walliapp C., Haehnle S., Mueller-Pillach F., Burghardt B.,
RA Iwamura T., Rutenberger M., Lerch M.M., Adler G., Gress T.M.;
RA "A novel transmembrane serine protease (TMPRSS3) overexpressed in
RA pancreatic cancer";
RA Cancer Res. 60:2602-2606(2000).
RN
RP SEQUENCE FROM N.A.
RP Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;
RT "MT-SP2, a novel type II membrane serine protease expressed in
RT trachea, colon, and small intestine: identification, cloning, and
RT chromosomal localization";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBS databases.
RN
RP SEQUENCE FROM N.A.
RP TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Udell T.B., Tothlyuk S., Carninci P., Prange C.,
RA Baeak S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Roark S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzley K.C., Hale S., Garcia A.M., Gay L., Hulyk S.W.,
RA Richards S., Morley K.C., Sodergren E.U., Lu X., Gibbs R.A.,
RA Pahay J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Green E.D., Dickson M.C.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywiński M.I., Skalska U., Smailus D.E.,
RA Schercher A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN
RL PROC. NATL. ACADEMY OF SCI. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: MAY PLAY A ROLE IN TUMOR PROGRESSION, METASTASIS
CC -1- FORMATION AND TUMOR INVASION.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN PANCREATIC, GASTRIC, COLORECTAL
CC AND AMPULLARY CANCER. VERY WEAK EXPRESSION IN NORMAL
CC GASTROINTESTINAL AND UROGENITAL TRACT.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: Contains 1 SRCR domain.
CC -1- SIMILARITY: Contains 1 LDL-receptor class A domain.
CC
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF179224; AAF74526.1; -
CC EMBL; AF216312; AAF31436.1; -
CC EMBL; BC011703; AAH11703.1; -
CC HSSP; P00763; IDPO.
CC
CC MEROFS; S01.034; -
CC
CC Genew; HGNC:11878; TMPRSS4.
CC MIM; 606565; -
CC GO; GO:0016021; C:Integral to membrane; NAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR002172; LD1_receptor_A.
CC InterPro; IPR001254; Ser_protease_Try.
CC InterPro; IPR001190; Src_receptor.
CC Pfam; PF00057; Id1_recept_a; 1.
CC Pfam; PF00059; Trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00192; LDla; 1.

```


DR	EMBL	U08837	AAA4026.1	-	
DR	PIR	A43090	AAA4035.1	-	
DR	PDB	1EKH	14-OCT-99		
DR	InterPro	IPR001314	Chymotrypsin.		
DR	InterPro	IPR000859	CUB_domain.		
DR	InterPro	IPR002172	LDL_receptor_A.		
DR	InterPro	IPR000938	MAM_domain.		
DR	InterPro	IPR001254	Set_protase_Try.		
DR	InterPro	IPR001190	Srcr_receptor.		
DR	Pfam	PF00431	CUB; 2.		
DR	Pfam	PF00057	Idl_recept_a; 2.		
DR	Pfam	PF00629	MAM; 1.		
DR	Pfam	PF01390	SEA; 1.		
DR	Pfam	PF00530	SRCR; 1.		
DR	PRINTS	PRO0089	Lypsin; 1.		
DR	PRINTS	PRO0722	CHYMOTRYPSIN.		
DR	SMART	SMO0261	LDLRECEPTOR.		
DR	SMART	SMO0442	CUB; 2.		
DR	SMART	SMO0192	LDLa; 2.		
DR	SMART	SMO0137	MAM; 1.		
DR	SMART	SMO0200	SEA; 1.		
DR	SMART	SMO0202	SR; 1.		
DR	SMART	SMO0180	Tryp_Spec; 1.		
DR	PROSITE	PS01180	CUB; 2.		
DR	PROSITE	PS01209	LDLRA_1; 2.		
DR	PROSITE	PS00068	LDLRA_2; 2.		
DR	PROSITE	PS00740	MAM_1; 1.		
DR	PROSITE	PS00060	MAM_2; 1.		
DR	PROSITE	PS00024	SEA; 1.		
DR	PROSITE	PS00420	SRCR_1; PAUSE-NEG.		
DR	PROSITE	PS0287	SRCR_2; 1.		
DR	PROSITE	PS0240	TRYPSIN_DOM; 1.		
DR	PROSITE	PS00134	TRYPSIN_HIS; 1.		
DR	PROSITE	PS00135	TRYPSIN_SER; 1.		
KW	Signal.anchor		Glycoprotein; Myristate; Hydroxylase; Serine protease; Zymogen; Transmembrane; Repeat; Alternative splicing; 3D-structure.		
KW	CHAIN	1	800		
FT	CHAIN	801	1035		
FT	CHAIN	1	18		
FT	TRANSMEM	19	47		
FT	DOMAIN	48	1035		
FT	DOMAIN	54	169		
FT	DOMAIN	197	238		
FT	DOMAIN	240	350		
FT	DOMAIN	358	520		
FT	DOMAIN	540	650		
FT	DOMAIN	657	787		
FT	DOMAIN	694	895		
FT	DOMAIN	801	1035		
FT	ACT_SITE	841	843		
FT	ACT_SITE	892	892		
FT	ACT_SITE	987	987		
FT	LIPID	2	2		
FT	DISULFID	199	212		
FT	DISULFID	206	225		
FT	DISULFID	219	236		
FT	DISULFID	659	671		
FT	DISULFID	666	684		
FT	DISULFID	678	693		
FT	DISULFID	788	912		
FT	DISULFID	826	842		
FT	DISULFID	926	993		
FT	DISULFID	957	972		
FT	DISULFID	983	1011		

FT	CARBOHYD	116	116	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	147	147	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	170	170	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	194	194	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	233	233	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	263	263	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	264	264	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	404	404	N-LINKED (GLCNAC . .)	(POTENTIAL)
PT	CARBOHYD	456	456	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	486	486	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	519	519	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	550	550	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	646	646	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	698	698	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	722	722	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	741	741	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	762	762	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	864	864	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	903	903	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	VASAPPLIC	965	965	N-LINKED (GLCNAC . .)	(POTENTIAL)
FT	VASAPPLIC	166	192	Missing (in isoform Short).	
FT	CONFLICT	808	808	R--Y (IN REF. 3).	
SQ	SEQUENCE	1035 AA;	114887 FM; E207970B08296E13 CRC64;		
 Query Match 34.8% Score 536.5; DB 1; Length 1035; Best Local Similarity 39.2%; Pred. No. 1.5e-40; Matches 111; Conservative 53; Mismatches 104; Indels 15; Gaps 6					
Oy	1 MKLNTSAGVVDIYKRLVHSDACSSEKAVSLRC--IAGCVNLSSROS-RIVGSEBALPCA	57			
Dd	755 VILATDPNGSL---LTSPQQCLEDSLIIAQCYKSCGGKLTVQEVSPRTVGSSDSREGA	811			
Oy	58 WPMOVSILAVQNVHVCGSIITPEWIVTAHCVEKEPLNFMHTAFAGILROSFMYGAGY	117			
Dd	812 MPVVVALYEDDDQVCASLVSRDWLSALHCVTGRMMESPKNKAIVLGIMANL---TSP	868			
Oy	118 QVR-----KVISHPVYSKTQNDIAAMKQKPLTRNDLYKPYCLPNFGMLQBPQLCMT	172			
Dd	869 QJSTRLLIDVYNPHNRKRNDIAMMELMKVNTVDYIOPICLPEENOVPPPRICSI	928			
Oy	173 SGWCATEEKGTISEVTLNAKVLLETQRCSRSRYVDNLITTPAMI CAGFLQGVDSCQGS	232			
Dd	929 ACGMALIVYGSTADVLQADVPLINSKKCCQQCKPEYN-ITENNVCAGYAGGVDSQCQS	987			
Oy	233 GGLVTSKRNIMWLIGDTSMGSGCAKAYRPGVGNMVFTDI	275			
Dd	988 GGELMCQENRMFLAGVTSFQYOCALFNREGVYARVPRTFWI	1030			
 RESULT 11					
ENTX_HUMAN	ID ENTX_HUMAN	STANDARD:	PRT:	1019 AA.	
AC	P98073;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Enteropeptidase precursor (EC 3.4.21.9) (Enterokinase).				
GN	PEPS7 OR ENTK.				
OS	Homo sapiens (Human).				
OC	Bakayocia, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Ductenum;				
RX	MEDLINE=95234679; PubMed=7718557;				
RA	Kilianovic Y., Vejle R.A., Doms-Jøeller H., Sadler J.E.;				
RT	"CDNA sequence and chromosomal localization of human enterokinas,				
RL	the proteolytic activator of trypsinogen.";				
RU	Biochemistry 34;4562-4568(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A., AND DISEASE-				

RA MEDLINE=21606074; PubMed=11719902;
 RA Holzinger A., Maiter E.M., Buck C., Meyerhofer P.U., Kappeler M.,
 RA Haworth J.C., Moroz S.P., Hadorn H.-B., Sadler J.E., Roehrer A.A.,
 RT "Mutations in the proenteropeptidase gene are the molecular cause of
 RT congenital enteropeptidase deficiency.";
 RL Am. J. Hum. Genet. 70:20-25(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Matanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Toroki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schlinnabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordle K., Horstischer K., Brandt P.,
 RA Schaefer M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
 RA Mehrezy S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L., Nizetic D., Francis F.,
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [4]
 RP SEQUENCE OF 749-1019 FROM N.A.
 RX TISSUE-Duodenum;
 RX MEDLINE=94329561; PubMed=8052624;
 RA Kitanoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.,
 RT enterokinase, the initiator of intestinal digestion, is a mosaic
 RT protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC CARBOXYPEPTIDASES, AND PROELASTASES.
 CC -1- CATALYTIC ACTIVITY: Selective cleavage of 6-Lys|-Ile-7 bond in
 CC trypsinogen. HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC SUBUNIT. HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -1- DISEASE: Defects in PRSS7 are a cause of enterokinase deficiency
 CC [MIM:262001]; a life-threatening intestinal malabsorption disorder
 CC characterized by diarrhea and failure to thrive.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -1- SIMILARITY: Contains 2 LDL-receptor class A domains.
 CC -1- SIMILARITY: Contains 2 CUB domains.
 CC -1- SIMILARITY: Contains 1 SEA domain.
 CC -1- SIMILARITY: Contains 1 SRCR domain.
 CC -1- SIMILARITY: Contains 1 MAM domain.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; Y19134; CAB65555.1; JOINED.
 DR EMBL; Y19135; CAB65555.1; JOINED.
 DR EMBL; Y19136; CAB65555.1; JOINED.
 DR EMBL; Y19137; CAB65555.1; JOINED.
 DR EMBL; Y19138; CAB65555.1; JOINED.
 DR EMBL; Y19139; CAB65555.1; JOINED.
 DR EMBL; Y19140; CAB65555.1; JOINED.
 DR EMBL; Y19141; CAB65555.1; JOINED.
 DR EMBL; Y19142; CAB65555.1; JOINED.
 DR EMBL; Y19143; CAB65555.1; JOINED.
 DR EMBL; Y19144; CAB65555.1; JOINED.
 DR EMBL; Y19145; CAB65555.1; JOINED.
 DR EMBL; Y19146; CAB65555.1; JOINED.
 DR EMBL; Y19147; CAB65555.1; JOINED.
 DR EMBL; Y19148; CAB65555.1; JOINED.
 DR EMBL; Y19149; CAB65555.1; JOINED.
 DR EMBL; Y19150; CAB65555.1; JOINED.
 DR EMBL; Y19151; CAB65555.1; JOINED.
 DR EMBL; Y19152; CAB65555.1; JOINED.
 DR EMBL; Y19153; CAB65555.1; JOINED.
 DR EMBL; Y19154; CAB65555.1; JOINED.
 DR EMBL; Y19155; CAB65555.1; JOINED.
 DR EMBL; Y19156; CAB65555.1; JOINED.
 DR EMBL; Y19157; CAB65555.1; JOINED.
 DR EMBL; Y19158; CAB65555.1; JOINED.
 DR EMBL; Y19159; CAB65555.1; JOINED.
 DR EMBL; Y19160; CAB65555.1; JOINED.
 DR EMBL; Y19161; CAB65555.1; JOINED.
 DR EMBL; Y19162; CAB65555.1; JOINED.
 DR EMBL; Y19163; CAB65555.1; JOINED.
 DR EMBL; Y19164; CAB65555.1; JOINED.
 DR EMBL; Y19165; CAB65555.1; JOINED.
 DR EMBL; Y19166; CAB65555.1; JOINED.
 DR EMBL; Y19167; CAB65555.1; JOINED.
 DR EMBL; Y19168; CAB65555.1; JOINED.
 DR EMBL; Y19169; CAB65555.1; JOINED.
 DR EMBL; Y19170; CAB65555.1; JOINED.
 DR EMBL; Y19171; CAB65555.1; JOINED.
 DR EMBL; Y19172; CAB65555.1; JOINED.
 DR EMBL; Y19173; CAB65555.1; JOINED.
 DR EMBL; Y19174; CAB65555.1; JOINED.
 DR EMBL; Y19175; CAB65555.1; JOINED.
 DR EMBL; Y19176; CAB65555.1; JOINED.
 DR EMBL; Y19177; CAB65555.1; JOINED.
 DR EMBL; Y19178; CAB65555.1; JOINED.
 DR EMBL; Y19179; CAB65555.1; JOINED.
 DR EMBL; Y19180; CAB65555.1; JOINED.
 DR EMBL; Y19181; CAB65555.1; JOINED.
 DR EMBL; Y19182; CAB65555.1; JOINED.
 DR EMBL; Y19183; CAB65555.1; JOINED.
 DR EMBL; Y19184; CAB65555.1; JOINED.
 DR EMBL; Y19185; CAB65555.1; JOINED.
 DR EMBL; Y19186; CAB65555.1; JOINED.
 DR EMBL; Y19187; CAB65555.1; JOINED.
 DR EMBL; Y19188; CAB65555.1; JOINED.
 DR EMBL; Y19189; CAB65555.1; JOINED.
 DR EMBL; Y19190; CAB65555.1; JOINED.
 DR EMBL; Y19191; CAB65555.1; JOINED.
 DR EMBL; Y19192; CAB65555.1; JOINED.
 DR EMBL; Y19193; CAB65555.1; JOINED.
 DR EMBL; Y19194; CAB65555.1; JOINED.
 DR EMBL; Y19195; CAB65555.1; JOINED.
 DR EMBL; Y19196; CAB65555.1; JOINED.
 DR EMBL; Y19197; CAB65555.1; JOINED.
 DR EMBL; Y19198; CAB65555.1; JOINED.
 DR EMBL; Y19199; CAB65555.1; JOINED.
 DR EMBL; Y19200; CAB65555.1; JOINED.
 DR EMBL; Y19201; CAB65555.1; JOINED.
 DR EMBL; Y19202; CAB65555.1; JOINED.
 DR EMBL; Y19203; CAB65555.1; JOINED.
 DR EMBL; Y19204; CAB65555.1; JOINED.
 DR EMBL; Y19205; CAB65555.1; JOINED.
 DR EMBL; Y19206; CAB65555.1; JOINED.
 DR EMBL; Y19207; CAB65555.1; JOINED.
 DR EMBL; Y19208; CAB65555.1; JOINED.
 DR EMBL; Y19209; CAB65555.1; JOINED.
 DR EMBL; Y19210; CAB65555.1; JOINED.
 DR EMBL; Y19211; CAB65555.1; JOINED.
 DR EMBL; Y19212; CAB65555.1; JOINED.
 DR EMBL; Y19213; CAB65555.1; JOINED.
 DR EMBL; Y19214; CAB65555.1; JOINED.
 DR EMBL; Y19215; CAB65555.1; JOINED.
 DR EMBL; Y19216; CAB65555.1; JOINED.
 DR EMBL; Y19217; CAB65555.1; JOINED.
 DR EMBL; Y19218; CAB65555.1; JOINED.
 DR EMBL; Y19219; CAB65555.1; JOINED.
 DR EMBL; Y19220; CAB65555.1; JOINED.
 DR EMBL; Y19221; CAB65555.1; JOINED.
 DR EMBL; Y19222; CAB65555.1; JOINED.
 DR EMBL; Y19223; CAB65555.1; JOINED.
 DR EMBL; Y19224; CAB65555.1; JOINED.
 DR EMBL; Y19225; CAB65555.1; JOINED.
 DR EMBL; Y19226; CAB65555.1; JOINED.
 DR EMBL; Y19227; CAB65555.1; JOINED.
 DR EMBL; Y19228; CAB65555.1; JOINED.
 DR EMBL; Y19229; CAB65555.1; JOINED.
 DR EMBL; Y19230; CAB65555.1; JOINED.
 DR EMBL; Y19231; CAB65555.1; JOINED.
 DR EMBL; Y19232; CAB65555.1; JOINED.
 DR EMBL; Y19233; CAB65555.1; JOINED.
 DR EMBL; Y19234; CAB65555.1; JOINED.
 DR EMBL; Y19235; CAB65555.1; JOINED.
 DR EMBL; Y19236; CAB65555.1; JOINED.
 DR EMBL; Y19237; CAB65555.1; JOINED.
 DR EMBL; Y19238; CAB65555.1; JOINED.
 DR EMBL; Y19239; CAB65555.1; JOINED.
 DR EMBL; Y19240; CAB65555.1; JOINED.
 DR EMBL; Y19241; CAB65555.1; JOINED.
 DR EMBL; Y19242; CAB65555.1; JOINED.
 DR EMBL; Y19243; CAB65555.1; JOINED.
 DR EMBL; Y19244; CAB65555.1; JOINED.
 DR EMBL; Y19245; CAB65555.1; JOINED.
 DR EMBL; Y19246; CAB65555.1; JOINED.
 DR EMBL; Y19247; CAB65555.1; JOINED.
 DR EMBL; Y19248; CAB65555.1; JOINED.
 DR EMBL; Y19249; CAB65555.1; JOINED.
 DR EMBL; Y19250; CAB65555.1; JOINED.
 DR EMBL; Y19251; CAB65555.1; JOINED.
 DR EMBL; Y19252; CAB65555.1; JOINED.
 DR EMBL; Y19253; CAB65555.1; JOINED.
 DR EMBL; Y19254; CAB65555.1; JOINED.
 DR EMBL; Y19255; CAB65555.1; JOINED.
 DR EMBL; Y19256; CAB65555.1; JOINED.
 DR EMBL; Y19257; CAB65555.1; JOINED.
 DR EMBL; Y19258; CAB65555.1; JOINED.
 DR EMBL; Y19259; CAB65555.1; JOINED.
 DR EMBL; Y19260; CAB65555.1; JOINED.
 DR EMBL; Y19261; CAB65555.1; JOINED.
 DR EMBL; Y19262; CAB65555.1; JOINED.
 DR EMBL; Y19263; CAB65555.1; JOINED.
 DR EMBL; Y19264; CAB65555.1; JOINED.
 DR EMBL; Y19265; CAB65555.1; JOINED.
 DR EMBL; Y19266; CAB65555.1; JOINED.
 DR EMBL; Y19267; CAB65555.1; JOINED.
 DR EMBL; Y19268; CAB65555.1; JOINED.
 DR EMBL; Y19269; CAB65555.1; JOINED.
 DR EMBL; Y19270; CAB65555.1; JOINED.
 DR EMBL; Y19271; CAB65555.1; JOINED.
 DR EMBL; Y19272; CAB65555.1; JOINED.
 DR EMBL; Y19273; CAB65555.1; JOINED.
 DR EMBL; Y19274; CAB65555.1; JOINED.
 DR EMBL; Y19275; CAB65555.1; JOINED.
 DR EMBL; Y19276; CAB65555.1; JOINED.
 DR EMBL; Y19277; CAB65555.1; JOINED.
 DR EMBL; Y19278; CAB65555.1; JOINED.
 DR EMBL; Y19279; CAB65555.1; JOINED.
 DR EMBL; Y19280; CAB65555.1; JOINED.
 DR EMBL; Y19281; CAB65555.1; JOINED.
 DR EMBL; Y19282; CAB65555.1; JOINED.
 DR EMBL; Y19283; CAB65555.1; JOINED.
 DR EMBL; Y19284; CAB65555.1; JOINED.
 DR EMBL; Y19285; CAB65555.1; JOINED.
 DR EMBL; Y19286; CAB65555.1; JOINED.
 DR EMBL; Y19287; CAB65555.1; JOINED.
 DR EMBL; Y19288; CAB65555.1; JOINED.
 DR EMBL; Y19289; CAB65555.1; JOINED.
 DR EMBL; Y19290; CAB65555.1; JOINED.
 DR EMBL; Y19291; CAB65555.1; JOINED.
 DR EMBL; Y19292; CAB65555.1; JOINED.
 DR EMBL; Y19293; CAB65555.1; JOINED.
 DR EMBL; Y19294; CAB65555.1; JOINED.
 DR EMBL; Y19295; CAB65555.1; JOINED.
 DR EMBL; Y19296; CAB65555.1; JOINED.
 DR EMBL; Y19297; CAB65555.1; JOINED.
 DR EMBL; Y19298; CAB65555.1; JOINED.
 DR EMBL; Y19299; CAB65555.1; JOINED.
 DR EMBL; Y19300; CAB65555.1; JOINED.
 DR EMBL; Y19301; CAB65555.1; JOINED.
 DR EMBL; Y19302; CAB65555.1; JOINED.
 DR EMBL; Y19303; CAB65555.1; JOINED.
 DR EMBL; Y19304; CAB65555.1; JOINED.
 DR EMBL; Y19305; CAB65555.1; JOINED.
 DR EMBL; Y19306; CAB65555.1; JOINED.
 DR EMBL; Y19307; CAB65555.1; JOINED.
 DR EMBL; Y19308; CAB65555.1; JOINED.
 DR EMBL; Y19309; CAB65555.1; JOINED.
 DR EMBL; Y19310; CAB65555.1; JOINED.
 DR EMBL; Y19311; CAB65555.1; JOINED.
 DR EMBL; Y19312; CAB65555.1; JOINED.
 DR EMBL; Y19313; CAB65555.1; JOINED.
 DR EMBL; Y19314; CAB65555.1; JOINED.
 DR EMBL; Y19315; CAB65555.1; JOINED.
 DR EMBL; Y19316; CAB65555.1; JOINED.
 DR EMBL; Y19317; CAB65555.1; JOINED.
 DR EMBL; Y19318; CAB65555.1; JOINED.
 DR EMBL; Y19319; CAB65555.1; JOINED.
 DR EMBL; Y19320; CAB65555.1; JOINED.
 DR EMBL; Y19321; CAB65555.1; JOINED.
 DR EMBL; Y19322; CAB65555.1; JOINED.
 DR EMBL; Y19323; CAB65555.1; JOINED.
 DR EMBL; Y19324; CAB65555.1; JOINED.
 DR EMBL; Y19325; CAB65555.1; JOINED.
 DR EMBL; Y19326; CAB65555.1; JOINED.
 DR EMBL; Y19327; CAB65555.1; JOINED.
 DR EMBL; Y19328; CAB65555.1; JOINED.
 DR EMBL; Y19329; CAB65555.1; JOINED.
 DR EMBL; Y19330; CAB65555.1; JOINED.
 DR EMBL; Y19331; CAB65555.1; JOINED.
 DR EMBL; Y19332; CAB65555.1; JOINED.
 DR EMBL; Y19333; CAB65555.1; JOINED.
 DR EMBL; Y19334; CAB65555.1; JOINED.
 DR EMBL; Y19335; CAB65555.1; JOINED.
 DR EMBL; Y19336; CAB65555.1; JOINED.
 DR EMBL; Y19337; CAB65555.1; JOINED.
 DR EMBL; Y19338; CAB65555.1; JOINED.
 DR EMBL; Y19339; CAB65555.1; JOINED.
 DR EMBL; Y19340; CAB65555.1; JOINED.
 DR EMBL; Y19341; CAB65555.1; JOINED.
 DR EMBL; Y19342; CAB65555.1; JOINED.
 DR EMBL; Y19343; CAB65555.1; JOINED.
 DR EMBL; Y19344; CAB65555.1; JOINED.
 DR EMBL; Y19345; CAB65555.1; JOINED.
 DR EMBL; Y19346; CAB65555.1; JOINED.
 DR EMBL; Y19347; CAB65555.1; JOINED.
 DR EMBL; Y19348; CAB65555.1; JOINED.
 DR EMBL; Y19349; CAB65555.1; JOINED.
 DR EMBL; Y19350; CAB65555.1; JOINED.
 DR EMBL; Y19351; CAB65555.1; JOINED.
 DR EMBL; Y19352; CAB65555.1; JOINED.
 DR EMBL; Y19353; CAB65555.1; JOINED.
 DR EMBL; Y19354; CAB65555.1; JOINED.
 DR EMBL; Y19355; CAB65555.1; JOINED.
 DR EMBL; Y19356; CAB65555.1; JOINED.
 DR EMBL; Y19357; CAB65555.1; JOINED.
 DR EMBL; Y19358; CAB65555.1; JOINED.
 DR EMBL; Y19359; CAB65555.1; JOINED.
 DR EMBL; Y19360; CAB65555.1; JOINED.
 DR EMBL; Y19361; CAB65555.1; JOINED.
 DR EMBL; Y19362; CAB65555.1; JOINED.
 DR EMBL; Y19363; CAB65555.1; JOINED.
 DR EMBL; Y19364; CAB65555.1; JOINED.
 DR EMBL; Y19365; CAB65555.1; JOINED.
 DR EMBL; Y19366; CAB65555.1; JOINED.
 DR EMBL; Y19367; CAB65555.1; JOINED.
 DR EMBL; Y19368; CAB65555.1; JOINED.
 DR EMBL; Y19369; CAB65555.1; JOINED.
 DR EMBL; Y19370; CAB65555.1; JOINED.
 DR EMBL; Y19371; CAB65555.1; JOINED.
 DR EMBL; Y19372; CAB65555.1; JOINED.
 DR EMBL; Y19373; CAB65555.1; JOINED.
 DR EMBL; Y19374; CAB65555.1; JOINED.
 DR EMBL; Y19375; CAB65555.1; JOINED.
 DR EMBL; Y19376; CAB65555.1; JOINED.
 DR EMBL; Y19377; CAB65555.1; JOINED.
 DR EMBL; Y19378; CAB65555.1; JOINED.
 DR EMBL; Y19379; CAB65555.1; JOINED.
 DR EMBL; Y19380; CAB65555.1; JOINED.
 DR EMBL; Y19381; CAB65555.1; JOINED.
 DR EMBL; Y19382; CAB65555.1; JOINED.
 DR EMBL; Y19383; CAB65555.1; JOINED.
 DR EMBL; Y19384; CAB65555.1; JOINED.
 DR EMBL; Y19385; CAB65555.1; JOINED.
 DR EMBL; Y19386; CAB65555.1; JOINED.
 DR EMBL; Y19387; CAB65555.1; JOINED.
 DR EMBL; Y19388; CAB65555.1; JOINED.
 DR EMBL; Y19389; CAB65555.1; JOINED.
 DR EMBL; Y19390; CAB65555.1; JOINED.
 DR EMBL; Y19391; CAB65555.1; JOINED.
 DR EMBL; Y19392; CAB65555.1; JOINED.
 DR EMBL; Y19393; CAB65555.1; JOINED.
 DR EMBL; Y19394; CAB65555.1; JOINED.
 DR EMBL; Y19395; CAB65555.1; JOINED.
 DR EMBL; Y19396; CAB65555.1; JOINED.
 DR EMBL; Y19397; CAB65555.1; JOINED.
 DR EMBL; Y19398; CAB65555.1; JOINED.
 DR EMBL; Y19399; CAB65555.1; JOINED.
 DR EMBL; Y19400; CAB65555.1; JOINED.
 DR EMBL; Y19401; CAB65555.1; JOINED.
 DR EMBL; Y19402; CAB65555.1; JOINED.
 DR EMBL; Y19403; CAB65555.1; JOINED.
 DR EMBL; Y19404; CAB65555.1; JOINED.
 DR EMBL; Y19405; CAB65555.1; JOINED.
 DR EMBL; Y19406; CAB65555.1; JOINED.
 DR EMBL; Y19407; CAB65555.1; JOINED.
 DR EMBL; Y19408; CAB65555.1; JOINED.
 DR EMBL; Y19409; CAB65555.1; JOINED.
 DR EMBL; Y19410; CAB65555.1; JOINED.
 DR EMBL; Y19411; CAB65555.1; JOINED.
 DR EMBL; Y19412; CAB65555.1; JOINED.
 DR EMBL; Y19413; CAB65555.1; JOINED.
 DR EMBL; Y19414; CAB65555.1; JOINED.
 DR EMBL; Y19415; CAB65555.1; JOINED.
 DR EMBL; Y19416; CAB65555.1; JOINED.
 DR EMBL; Y19417; CAB65555.1; JOINED.
 DR EMBL; Y19418; CAB65555.1; JOINED.
 DR EMBL; Y19419; CAB65555.1; JOINED.
 DR EMBL; Y19420; CAB65555.1; JOINED.
 DR EMBL; Y19421; CAB65555.1; JOINED.
 DR EMBL; Y19422; CAB65555.1; JOINED.
 DR EMBL; Y19423; CAB65555.1; JOINED.
 DR EMBL; Y19424; CAB65555.1; JOINED.
 DR EMBL; Y19425; CAB65555.1; JOINED.
 DR EMBL; Y19426; CAB65555.1; JOINED.
 DR EMBL; Y19427; CAB65555.1; JOINED.
 DR EMBL; Y19428; CAB65555.1; JOINED.
 DR EMBL; Y19429; CAB65555.1; JOINED.
 DR EMBL; Y19430; CAB65555.1; JOINED.
 DR EMBL; Y19431; CAB65555.1; JOINED.
 DR EMBL; Y19432; CAB65555.1; JOINED.
 DR EMBL; Y19433; CAB65555.1; JOINED.
 DR EMBL; Y19434; CAB65555.1; JOINED.
 DR EMBL; Y19435; CAB65555.1; JOINED.
 DR EMBL; Y19436; CAB65555.1; JOINED.
 DR EMBL; Y19437; CAB65555.1; JOINED.
 DR EMBL; Y19438; CAB65555.1; JOINED.
 DR EMBL; Y19439; CAB65555.1; JOINED.
 DR EMBL; Y19440; CAB65555.1; JOINED.
 DR EMBL; Y19441; CAB65555.1; JOINED.
 DR EMBL; Y19442; CAB65555.1; JOINED.
 DR EMBL; Y19443; CAB65555.1; JOINED.
 DR EMBL; Y19444; CAB65555.1; JOINED.
 DR EMBL; Y19445; CAB65555.1; JOINED.
 DR EMBL; Y19446; CAB65555.1; JOINED.
 DR EMBL; Y19447; CAB65555.1; JOINED.
 DR EMBL; Y19448; CAB65555.1; JOINED.
 DR EMBL; Y19449; CAB65555.1; JOINED.
 DR EMBL; Y19450; CAB65555.1; JOINED.
 DR EMBL; Y19451; CAB65555.1; JOINED.
 DR EMBL; Y19452; CAB65555.1; JOINED.
 DR EMBL; Y19453; CAB65555.1; JOINED.
 DR EMBL; Y19454; CAB65555.1; JOINED.
 DR EMBL; Y19455; CAB65555.1; JOINED.
 DR EMBL; Y19456; CAB65555.1; JOINED.
 DR EMBL; Y19457; CAB65555.1; JOINED.
 DR EMBL; Y19458; CAB65555.1; JOINED.
 DR EMBL; Y19459; CAB65555.1; JOINED.
 DR EMBL; Y19460; CAB65555.1; JOINED.
 DR EMBL; Y19461; CAB65555.1; JOINED.
 DR EMBL; Y19462; CAB65555.1; JOINED.
 DR EMBL; Y19463; CAB65555.1; JOINED.
 DR EMBL; Y19464; CAB65555.1; JOINED.
 DR EMBL; Y19465; CAB65555.1; JOINED.
 DR EMBL; Y19466; CAB65555.1; JOINED.
 DR EMBL; Y19467; CAB65555.1; JOINED.
 DR EMBL; Y19468; CAB65555.1; JOINED.
 DR EMBL; Y19469; CAB65555.1; JOINED.
 DR EMBL; Y19470; CAB65555.1; JOINED.
 DR EMBL; Y19471; CAB65555.1; JOINED.
 DR EMBL; Y19472; CAB65555.1; JOINED.
 DR EMBL; Y19473; CAB65555.1; JOINED.
 DR EMBL; Y19474; CAB65555.1; JOINED.
 DR EMBL; Y19475; CAB65555.1; JOINED.
 DR EMBL; Y19476; CAB65555.1; JOINED.
 DR EMBL; Y19477; CAB65555.1; JOINED.
 DR EMBL; Y19478; CAB65555.1; JOINED.
 DR EMBL; Y19479; CAB65555.1; JOINED.
 DR EMBL; Y19480; CAB65555.1; JOINED.
 DR EMBL; Y19481; CAB65555.1; JOINED.
 DR EMBL; Y19482; CAB65555.1; JOINED.
 DR EMBL; Y19483; CAB65555.1; JOINED.
 DR EMBL; Y19484; CAB65555.1; JOINED.
 DR EMBL; Y19485; CAB65555.1; JOINED.
 DR EMBL; Y19486; CAB65555.1; JOINED.
 DR EMBL; Y19487; CAB65555.1; JOINED.
 DR EMBL; Y19488; CAB65555.1; JOINED.
 DR EMBL; Y19489; CAB65555.1; JOINED.
 DR EMBL; Y19490; CAB65555.1; JOINED.
 DR EMBL; Y19491; CAB65555.1; JOINED.
 DR EMBL; Y19492; CAB65555.1; JOINED.
 DR EMBL; Y19493; CAB65555.1; JOINED.
 DR EMBL; Y19494; CAB65555.1; JOINED.
 DR EMBL; Y19495; CAB65555.1; JOINED.
 DR EMBL; Y19496; CAB65555.1; JOINED.
 DR EMBL; Y19497; CAB65555.1; JOINED.
 DR EMBL; Y19498; CAB65555.1; JOINED.
 DR EMBL; Y19499; CAB65555.1; JOINED.
 DR EMBL; Y19500; CAB65555.1; JOINED.
 DR EMBL; Y19501; CAB65555.1; JOINED.
 DR EMBL; Y19502; CAB65555.1; JOINED.
 DR EMBL; Y19503; CAB65555.1; JOINED.
 DR EMBL; Y19504; CAB65555.1; JOINED.
 DR EMBL; Y19505; CAB65555.1; JOINED.
 DR EMBL; Y19506; CAB65555.1; JOINED.
 DR EMBL; Y19507; CAB65555.1; JOINED.
 DR EMBL; Y19508; CAB65555.1; JOINED.
 DR EMBL; Y19509; CAB65555.1; JOINED.
 DR EMBL; Y19510; CAB65555.1; JOINED.
 DR EMBL; Y19511; CAB65555.1; JOINED.
 DR EMBL; Y19512; CAB65555.1; JOINED.
 DR EMBL; Y19513; CAB65555.1; JOINED.
 DR EMBL; Y19514; CAB65555.1; JOINED.
 DR EMBL; Y19515; CAB65555.1; JOINED.
 DR EMBL; Y19516; CAB65555.1; JOINED.
 DR EMBL; Y19517; CAB65555.1; JOIN

FT DISULFID 191 210 BY SIMILARITY.
 FT DISULFID 204 221 BY SIMILARITY.
 FT DISULFID 643 655 BY SIMILARITY.
 FT DISULFID 650 668 BY SIMILARITY.
 FT DISULFID 662 677 BY SIMILARITY.
 FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 810 826 BY SIMILARITY.
 FT DISULFID 910 977 BY SIMILARITY.
 FT DISULFID 941 956 BY SIMILARITY.
 FT DISULFID 967 995 BY SIMILARITY.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 338 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 134 134 Q -> B (IN REF. 3).
 FT CONFLICT 732 732 S -> P (IN REF. 3).
 FT CONFLICT 754 771 SOOGLDLSIRLCNHNKS -> RRRAKNEIDALSPITLIA
 (IN REF. 3).
 SQ SEQUENCE 1019 AA, 112923 MW, B6AA245F6M4A563 CRC64;
 Query Match 34.3%; Score 528.5; DB 1; Length 1019;
 Best Local Similarity 38.6%; Pred. No. 7.5e-40;
 Matches 108; Conservative 56; Mismatches 107; Indels 9; Gaps 5;
 QY 1 MKLNTSGANDVYIKLYHSDACSSKAVSLRC--IACGVNINSSR--GSRIVGSSALPGA 57
 DB 739 VKLNTNP---DGLILITPBOOCLQDSLILQCNHNSCGKRLAADITPKIVGSSNAEGA 755
 QY 58 WPNVQVLAHONVAVCGSITTEBMTVAHCEKPLNPMEMTAPAGILRGSENF--YGA 115
 DB 796 WPNVAVLYYGRLLCASLVSDDMLVSAHCVYGNLSPSKWTALLIGLHMKSNLITSPQV 855
 QY 116 GYGVKVIKSHPNVDSKTKQNDIALMKLOKELFNDLVPCLEPNPGMLQBOELCWISGM 175
 DB 856 PRILDEIVINPHNRRKQNDIAMHLEKRYVITYIPICLPBENOVPEFGNCSIAGM 915
 QY 176 GATEBKGTSEVINAQVLLIFQTSKNSRYVDNLITPAMI CAGFLQGVNDS CGGDSGGX 235
 DB 916 GLVYVGGTANILQEDVPLTNSRCQQQMPHYN-ITENMI CAGYEGLDSCGDSGGP 974
 QY 236 LVTSKKNIMWLIGDTSWGSQCAKAYRPGYGVNMFDTMI 275
 DB 975 LMCQENRRFPLAGVTSFGYKCALPNRPGYVARSFTEMI 1014
 RESULT 12
 KAL_MOUSE STANDARD; PRT; 638 AA.
 ID KAL_MOUSE
 AC P26262;
 DT 01-MAY-1992 (Rel. 22, Last Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Plasma kallikrein precursor (BC 3.4.21.34) (Plasma prekallikrein)
 DE (kininogenin) (Fletcher factor).
 GN KLK3 OR KLK3 OR PK.
 OS Mus musculus (Mouse).
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;

BN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=BAUB/c; TISSUE=Liver;
 RX MEDLINE=91090844; PubMed=2264928;
 RA Seidah N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
 RA Brachpala L., Rochemont J., Mbikay M., Chretien M.,
 RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
 RT and comparison of protein and mRNA levels among species".
 RL DNA Cell Biol. 9:737-748(1990).
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -1- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Xaa and Lys-|-Xaa
 CC bonds, including Lys-|-Arg and Arg-|-Ser bonds in (human)
 CC kininogen to release bradykinin.
 CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMA KALLIKREIN
 CC SUBFAMILY.
 CC -1- SIMILARITY: Contains 4 apple domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M58588; AAA63393.1; -;
 CC PIR; A36557; KOMSPL.
 CC HSSP; P00750; IRTP.
 CC MEROPS; S01.212; -;
 CC MGD; MGI:102849; Klkb1.
 CC Interpro; IPR000177; Apple.
 CC Interpro; IPR003114; Chymotrypsin.
 CC Interpro; IPR003014; PAN.
 CC Interpro; IPR001254; Ser_protease_Try.
 CC Pfam; PF00024; PAN; 4.
 CC DR Pfam; PF00089; trypsin; 1.
 CC DR PRINTS; PRO0005; APPLEDOMAIN.
 CC DR PRINTS; PRO0722; CHYMOTRYPSIN.
 CC SMART; SM00223; APPLE; 4.
 CC SMART; SM00020; TRY; SPC; 1.
 CC PROSITE; PS00495; APPLE; 4.
 CC PROSITE; PS0240; TRYPSIN_DOM; 1.
 CC PROSITE; PS00134; TRYPSIN_HIS; 1.
 CC PROSITE; PS00135; TRYPSIN_SRR; 1.
 CC Hydrolysis: Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 CC Phospholysis: Blood coagulation; Inflammatory response; Liver;
 CC Repeat.
 CC KX
 CC Repeat.
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 390
 CC FT CHAIN 391 638
 CC FT DOMAIN 20 105
 CC FT DOMAIN 110 195
 CC FT DOMAIN 200 285
 CC FT DOMAIN 291 376
 CC FT DOMAIN 389 621
 CC FT CARBOHYD 127 127
 CC FT CARBOHYD 308 308
 CC FT CARBOHYD 396 396
 CC FT CARBOHYD 453 453
 CC FT CARBOHYD 494 494
 CC FT ACT_SITE 434 434
 CC FT ACT_SITE 483 483
 CC FT ACT_SITE 578 578
 CC FT DISULFID 21 104
 CC FT DISULFID 47 77
 CC PLASMA KALLIKREIN HEAVY CHAIN.
 CC PLASMA KALLIKREIN LIGHT CHAIN.
 CC APPLE 1.
 CC APPLE 2.
 CC APPLE 3.
 CC APPLE 4.
 CC SERINE PROTEASE.
 CC N-LINKED (GLCNAC. . .) (PROBABLE).
 CC N-LINKED (GLCNAC. . .) (PROBABLE).
 CC N-LINKED (GLCNAC. . .) (PROBABLE).
 CC O-LINKED (PROBABLE).
 CC N-LINKED (GLCNAC. . .) (PROBABLE).
 CC CATALYTIC ACTIVITY.
 CC CHARGE RELAY SYSTEM.
 CC CHARGE RELAY SYSTEM.
 CC CHARGE RELAY SYSTEM.
 CC BY SIMILARITY.
 CC BY SIMILARITY.

FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 111 194 BY SIMILARITY.
 FT DISULFID 137 166 BY SIMILARITY.
 FT DISULFID 141 147 BY SIMILARITY.
 FT DISULFID 201 284 BY SIMILARITY.
 FT DISULFID 227 256 BY SIMILARITY.
 FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 318 347 BY SIMILARITY.
 FT DISULFID 322 328 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 383 503 BY SIMILARITY.
 FT DISULFID 419 435 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;
 Query Match 34.1%; Score 526.5; DB 1; Length 638;
 Best Local Similarity 39.1%; Pseq No. 6.7e-40;
 Matches 106; Conservative 46; Mismatches 96; Indels 23; Gaps 5;
 QY 15 KYHSDACSKAVVSLRCLACGVNLSRSRRTVGSALPAPMPQVSLHVN--HY 71
 DB 376 KLVSDPDCCTTKI-----NARIVGTNALSIGMPQVSLQVTLVSGQTHL 418
 QY 72 CGGSITPEMITVTAACVCEKPLNPMHTAPAGILROSFMYGA-GYQVEKVISHPYDS 130
 DB 419 CGGSITGRMVLTAAHCPD-GIPYDWRITGILSLSEIRKPSRRIKELIHQEVK 477
 QY 131 KTRNDIALMKTQKPLTENDLVYVCLPDPGMLQPROLCISGKATBEKKTSLVLA 190
 DB 478 SEGNVDIAIKIQPLNTYTPQKPCLPSPKADVTWITYNQVWYMGYTKQGEFQNILOK 537
 QY 191 AKVLEIFORCNSRYVDNLITPAMICAFLOGVNDSCEGSGGLVTSKNINIMLIGDT 250
 DB 538 ATPLPVEECQCKK-RDYVINKMICAGYKEGEGDACKGSGGLVYCKSGKMQVLGVT 596
 QY 251 SWGSGCAKAYRPGYVGNVWTFDMYRQRA 281
 DB 597 SWEGGCGRQDPGYVTKYSEVMDILKRTQS 627
 RESULT 13
 HEP5_HUMAN STANDARD; PRT; 417 AA.
 ID P05961;
 AC 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Serine protease hepsin (EC 3.4.21.-) (Transmembrane protease, serine
 1).
 GN HEP OR TMPRSS1.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=86209431; PubMed=2835076;
 RA Leytus S.P., Leob K.R., Hagen F.S., Kurachi K., Davie E.W.;
 RT "A novel trypsin-like serine protease (hepsin) with a putative
 transmembrane domain expressed by human liver and hepatoma cells";
 RL Biochemistry 27:1067-1074(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancereas, and Spleen;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schejter C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schneez T.B.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mulaly S.J.,
 RA Bosak S.A., McKernan P.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalon D.R., Murny D.M., Sodergren B.J., Lu X., Gibbs K.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=91358502; PubMed=1885621;
 RA Tsuji A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,
 RA Chou S.H., Kurachi K.;
 RT "hepsin, a cell membrane-associated protease. Characterization,
 RT tissue distribution, and gene localization.";
 RL J. Biol. Chem. 266:16948-16953(1991).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=9348237; PubMed=8346233;
 RA Torres-Rosado A., O'Shea K.S., Tsuji A., Chou S.H., Kurachi K.;
 RT "hepsin, a putative cell-surface serine protease, is required for
 RT mammalian cell growth.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).
 CC -1- FUNCTION: Plays an essential role in cell growth and maintenance
 CC of cell morphology.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: Present in most tissues, with the highest
 CC level in liver.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC -----
 CC EMBL; M18930; AAA36013.1; -;
 CC EMBL; X07732; CA300558.1; -;
 CC EMBL; X07002; CA300558.1; -;
 CC EMBL; BC025716; AA25716.1; -;
 CC PIR; S00845; S00845.
 CC HSSP; P00763; IDPO.
 CC MEROPS; S01.224; -;
 CC GeneW; HGNC:5155; HPN.
 CC MIM; 142440; -;
 CC GO; GO:0005887; C:integral to plasma membrane; TAS.
 CC GO; GO:0008151; P:cell growth and/or maintenance; TAS.
 CC InterPro; IPR001314; Chymotrypsin.
 CC Pfam; P00089; trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC SMART; SM0020; TRYP_SPC; 1.
 CC PROSITE; PS50240; TRYPsin_DOM; 1.
 CC PROSITE; PS00134; TRYPsin_HIS; 1.
 CC PROSITE; PS00135; TRYPsin_SER; 1.
 CC KX Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
 CC SRRIN PROTEASE HEP5IN, NON-CATALYTIC
 CC CHAIN 162
 CC CHAIN 163 417
 CC DOMAIN 1 17
 CC TRANSMEM 18 44
 CC CYTOPLASMIC (POTENTIAL)
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL)

FT DOMAIN 45 417 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 163 417 SERINE PROTEASE.
 FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 153 277 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 188 204 BY SIMILARITY.
 FT DISULFID 322 338 BY SIMILARITY.
 FT DISULFID 349 381 BY SIMILARITY.
 FT CARBOHYD 112 112 N-LINKED (GLUCNA... (POTENTIAL).
 SO SEQUENCE 417 AA, 45011 MW, B2086PF661H551D7 CR664;
 Query Match 33.7%; Score 520.5; DB 1; Length 417;
 Best Local Similarity 38.7%; Pred. No. 1.4e-39;
 Matches 104; Conservative 43; Mismatches 109; Indels 13; Gaps 4;
 QY 22 CSSAAVSLACIACGVNLSRSRIVGSGSALPGANPQVSLHVNHHVGGSTITPEW 81
 DB 140 CPGREFLAICQDCG--RRKLPRDRIVGRDTSLSGMPQVSLRYDGAHLGGSLSGDW 197
 QY 82 IYTAACVCEKPLNNPMTAFAGILRQSFMYGAGYQYEVKYSHPNY-----DSKTKNN 135
 DB 198 VLPAAHCFEPERNRVLSRMRYFAGAVAGVQASP-HGLQGVQAVYHGGVLPFRDPRSRNSN 256
 QY 136 DILMLQKRLTRNDLVKPCPLNPGMLQPEOLCRISGKATBEKKTSEYLAIAKYL 195
 DB 257 DIALVHSSPLRLEYIQPCVLPAAQALVDKICLVYSGMNTQYVGOAGVLOEARVPI 316
 QY 196 IETORCSRYVYDNLITPAMICAGFLQANVDSQDSGGKLYL---TSKNNIMVIGDTS 251
 DB 317 ISNDVCGADPFYQNOIKPKMFCAGYEGGIDACQDSGSPFCESISRTPRMLCGIVS 376
 QY 252 WSGGCAKARHFGYGVGVNVMTMTYKQMR 280
 DB 377 WGTGCAIAQKPGVYTVSDPFRWIFQAIK 405
 RESULT 14
 FALL_HUMAN STANDARD; PRT; 625 AA.
 AC P03951; O9Y495;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Coagulation factor XI precursor (EC 3.4.21.27) (Plasma thromboplastin
 DE antecedent) (PFA) (FXI).
 GN PII.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=8624360; PubMed=3636155;
 RA Fujikawa K., Chung D.W., Hendrickson L.E., Davie E.W.;
 RT "Amino acid sequence of human factor XI, a blood coagulation factor
 RT with four tandem repeats that are highly homologous with plasma
 RT prekallikrein";
 RL Biochemistry 25:2417-2424(1986).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=88107663; PubMed=2827746;
 RA Asakai R., Davie E.W., Chung D.W.;
 RT "Organization of the gene for human factor XI";
 RL Biochemistry 26:7221-7228(1987).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=9826306; PubMed=9593722;
 RA Hau T.-C., Shore S.K., Seshemma T., Bagasra O., Walsh P.N.;
 RT "Molecular cloning of platelet factor XI, an alternative splicing
 RT product of the plasma factor XI gene";
 RL J. Biol. Chem. 273:13787-13793(1998).
 RN [4]

RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=91152017; PubMed=1998667;
 RA McMillen B.A., Fujikawa K., Davie E.W.;
 RT "Location of the disulfide bonds in human coagulation factor XI: the
 RT presence of tandem apple domains";
 RL Biochemistry 30:2056-2060(1991).
 RN [5]
 RP VARIANT LBU-301.
 RX MEDLINE=90046656; PubMed=2813350;
 RA Asakai R., Chung D.W., Ratnoff O.D., Davie E.W.;
 RT "Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi
 RT Jews is a bleeding disorder that can result from three types of point
 RT mutations";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).
 RN [6]
 RP VARIANT LBU-301.
 RX MEDLINE=92190478; PubMed=1547342;
 RA Meljers J.C., Davie E.W., Chung D.W.;
 RT "Expression of human blood coagulation factor XI: characterization of
 RT the defect in factor XI type III deficiency";
 RL Blood 79:1435-1440(1992).
 RN [7]
 RP VARIANTS HIS-34, PRO-320, ILE-322 AND LYS-341.
 RX MEDLINE=95195217; PubMed=7886672;
 RA Pugh R.E., McVey J.H., Tuddenham E.G., Hancock J.F.;
 RT "Six point mutations that cause factor XI deficiency";
 RL Blood 85:1509-1516(1995).
 RN [8]
 RP VARIANT VAL-460.
 RX MEDLINE=95399291; PubMed=7669672;
 RA Imanaka Y., Lai K., Nishimura T., Bolton-Maggs P.H., Tuddenham E.G.,
 RA McVey J.H.;
 RT "Identification of two novel mutations in non-Jewish factor XI
 RT deficiency";
 RL Br. J. Haematol. 90:916-920(1995).
 RN [9]
 RP VARIANT ASN-404.
 RX MEDLINE=98062240; PubMed=9401068;
 RA Wisnighausen B., Reischer A., Oddoux C., Oestreicher H., Nardi M.,
 RA Karpatkin M.;
 RT "Severe factor XI deficiency in an Arab family associated with a novel
 RT mutation in exon 11";
 RL Br. J. Haematol. 99:575-577(1997).
 RN [10]
 RP VARIANTS ARG-244 AND ASN-266.
 RX MEDLINE=99005359; PubMed=9787166;
 RA Martincic D., Zimmerman S.A., Ware R.E., Sun M.-F., Whitlock J.A.,
 RA Gallant D.;
 RT "Identification of mutations and polymorphisms in the factor XI genes
 RT of an African American family by dideoxyfingerprinting";
 RL Blood 92:3309-3317(1998).
 RN [11]
 RP ERRATUM.
 RA Martincic D., Zimmerman S.A., Ware R.E., Sun M.F., Whitlock J.A.,
 RA Gallant D.;
 RT Blood 93:1786-1786(1999).
 RN [12]
 RP VARIANT CYS-246.
 RX MEDLINE=99150037; PubMed=10027710;
 RA Albag A., Mitchell M., Sethi M., Rahman S., Flynn G., Boulton P.,
 RA Caeno G., Smith M., Savidge G.;
 RT "Identification of a novel mutation in a non-Jewish factor XI
 RT deficient kindred";
 RL Br. J. Haematol. 104:44-49(1999).
 RN [13]
 RP VARIANTS CYS-326; VAL-430 AND ARG-594.
 RX MEDLINE=20074632; PubMed=10606881;
 RA Mitchell M., Cutler J., Thompson S., Moore G., Jenkins Ap Rees E.,
 RA Smith M., Savidge G., Albag A.;
 RT "Heterozygous factor XI deficiency associated with three novel
 RT mutations";
 RL Br. J. Haematol. 107:763-765(1999).
 RN [14]

RA VARIANTS LBU-66; ARG-244; PHE-308 AND PHE-339.
RX MEDLINE=99318093; PubMed=10391209;
RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardle K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemes J., Ziaagra L.,
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA Lander E.S.;
RT "Characterization of single-nucleotide polymorphisms in coding regions
RT of human genes.";
RT Nat. Genet. 22:231-238 (1999).
RN (15)
RP ERRATUM.
RA Cargill M., Altschuler D., Ireland J., Sklar P., Ardle K., Patil N.,
RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemes J., Ziaagra L.,
RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
RA Lander E.S.;
RL Nat. Genet. 23:373-373 (1999).
CC -1- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC
CC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Arg-|-ala and Arg-|-val
CC bonds in factor IX to form factor IXa.
CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED. AFTER ACTIVATION THE HEAVY
CC AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P03951-1; Sequence=Displayed;
CC Name=2; Synonyms=Platelet;
CC IsoId=P03951-2; Sequence=VSP 005388;
CC -1- TISSUE SPECIFICITY: Isoform 2 is produced by platelets and
CC megakaryocytes but absent from other blood cells.
CC -1- PTM: ACTIVATED BY FACTOR XIa (OR XII), WHICH CLEAVES EACH
CC POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE
CC ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH
CC MOLECULAR WEIGHT (HMW) KININOGEN.
CC -1- DISEASE: Defects in F11 are a cause of a blood coagulation
CC abnormality (hereditary syndrome) occurring in high frequency in
CC Ashkenazi Jews. F11-deficient patients are prone to excessive
CC bleeding after haemostatic challenge.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMA KALLIKREIN
CC SUBFAMILY.
CC -1- SIMILARITY: Contains 4 apple domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).

DR GO; GO:0003603; F:Blood coagulation factor IX activity; NAS.
DR GO; GO:0003605; F:Blood coagulation factor XI activity; TAS.
DR GO; GO:0007996; P:Blood coagulation; NAS.
DR InterPro: IPR000177; Apple.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR003014; PAN.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00024; PAN; 4.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00005; APPLIEDMAIN.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00223; APPLR; 4.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS00495; APPLR; 4.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Glycoprotein; Plasma; Blood coagulation;
KW Repeat; Signal; Alternative splicing; Disease mutation; Polymorphism.
FT SIGNAL 1 18
Query Match 33.7%; Score 520; DB 1; Length 625;
Best Local Similarity 41.8%; Pred. No. 2.5e-39;
Matches 102; Conservative 43; Mismatches 85; Indels 14; Gaps 5;
QY 46 RIVGESALPGAMPQVSLVQNV---HVCGSITPEWITVAHCVKEPLNPMPTAF 102
DB 387 RIVGASASVGRKPMQVTLHTSPQRRLCGSITIGMILTANHC-YGVESKILRV 445
QY 103 AGILRQSFM---FYGAGYQVEKYSHPNTDSKNDIAMLQKPLFNDLVKVC 157
DB 446 SGLINSEIKEDTSPFG---VQRIHIDQYKMESEGYDALMLKLTFTVYTSORICL 501
QY 158 PNRGPMALQEQELCMISGKATEKSTSEVLAARVLLIETQRCSRYVDNLTTPMIC 217
DB 502 PNRGDMNVITTCWYGVGRKRLRKIKNTLQKAKIPLVTMBEQKRY-RGKHTHMC 560
QY 218 AGPLQGVNDSQGDSSGKLVYSKNNIMLIGTYSKSGSCAAVPGYGVNPTDIYR 277
DB 561 AGREGGKACKCGDSGGLSCNHEVHVLVGTISWEGCAQRRPGYTVNVEVDWILE 620
QY 278 QMR 281
DB 621 KTDQ 624
RESULT 15
HATT_HUMAN STANDARD; PRT; 418 AA.
AC 060235;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
GN Hatt.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88234382; PubMed=9565616;
RA Yamaoka K., Maeda K.-I., Ogawa H., Takagi K., Umehara N., Yamaoka S.;
RT "Cloning and characterization of the cDNA for human airway trypsin-
RT like protease.";
RL J. Biol. Chem. 273:11895-11901 (1998).
RN [2]
RP SEQUENCE OF 187-206, AND CHARACTERIZATION.
RX MEDLINE=97224034; PubMed=9070615;
RA Yamaoka K., Ohnishi T., Kawano S., Tsuchihashi S., Ogawara M.,
RA Maeda K.-I., Yamaoka K., Takahashi M., Sano T.;
RT "Purification, characterization, and localization of a novel
RT trypsin-like protease found in the human airway.";

Am. J. Respir. Cell Mol. Biol. 16:300-308(1997).

CC -1- FUNCTION: May play some biological role in the host defense system
 CC on the mucous membrane independently of or in cooperation with
 CC other substances in airway mucous or bronchial secretions.
 CC -1- CATALYTIC ACTIVITY: Preferentially cleaves the C-terminal side of
 CC arginine residues at the P1 position of certain peptides, cleaving
 CC Arg-Phe-Ser-Arg-4-methylcoumaryl-7-amide most efficiently and
 CC having an optimum pH of 8.6 with this substrate.
 CC -1- ENZYME REGULATION: Strongly inhibited by diisopropyl
 CC fluorophosphate, leupeptin, antipain, apricotinin, and soybean
 CC trypsin inhibitor, but hardly inhibited by secretory leukocyte
 CC protease inhibitor at 10 microm.

CC -1- SUBUNIT: Monomer.

CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN; ACTIVATED BY
 CC CLEAVAGE AND SECRETED.

CC -1- TISSUE SPECIFICITY: Located in the cells of the submucosal serous
 CC glands of the bronchi and trachea.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC -1- SIMILARITY: Contains 1 SEA domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AB002134; BAA28691.1; -

DR HSSP: P00750; 1RTP.

DR MEROPS: S01.301; -

DR MIM: 605369; -

DR GO: GO:0005576; C:extracellular; TNS.

DR GO: GO:0005887; C:integral to plasma membrane; TNS.

DR GO: GO:0008233; F:peptidase activity; TNS.

DR GO: GO:0007585; P:respiratory gaseous exchange; TNS.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR000082; SEA domain.

DR InterPro: IPR001254; Ser_protease_Try.

DR Pfam: PF01390; SEA; 1.

DR Pfam: PR00089; trypsin; 1.

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR SMART: SM00200; SEA; 1.

DR SMART: PS50024; TRYP_SPC; 1.

DR PROSITE: PS50240; TRYPIN_DOM; 1.

DR PROSITE: PS00134; TRYPIN_HIS; 1.

DR PROSITE: PS00135; TRYPIN_SER; 1.

KW Hydroxylase, Serine protease, Transmembrane, Signal-anchor, Zymogen,
 KW Glycoprotein.

FT CHAIN 1 186 AIRWAY TRYPSIN-LIKE PROTEASE, NON-
 FT CATALYTIC CHAIN.

FT CHAIN 187 418 AIRWAY TRYPSIN-LIKE PROTEASE, CATALYTIC
 FT CHAIN.

FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 21 41 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).

FT DOMAIN 42 418 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 44 164 SEA.

FT ACT_SITE 187 417 SERINE PROTEASE.

FT ACT_SITE 227 417 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 272 417 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 368 417 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT DISULFID 173 292 INTERCHAIN (BY SIMILARITY).

FT DISULFID 212 228 BY SIMILARITY.

FT DISULFID 337 353 BY SIMILARITY.

FT DISULFID 364 393 BY SIMILARITY.

FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).

SO SEQUENCE 418 AA; 46263 MW; F4BC1DB020CPBBD0 CRC64;

Query Match 33.6%; Score 518.5; DB 1; Length 418;
 Best Local Similarity 39.4%; Pred No. 2,2e-39;
 Matches 110; Conservative 43; Mismatches 117; Indels 9; Gaps 6;

Qy 3 INTSAGNDVIYKLYHSDACSSRAVSLRCIACGV--NUNSSROSRIVGSGSALPKGAMPW 60
 Db 143 LANS-GNLEINPST-ETITSLDAAANWMLINECGAPDILITLSEGIILCGTRAEKSWPW 200

Qy 61 QVSLHVNQVHVCGGSIITPEMIVTAHCVCKEPLNNPMHTAIPAGILROSPFMYGAGQYE 120
 Db 201 QVSLFNNAHCGGSLINNMILPAACF-RKSNINRDIATSGI--STTPKILMRNR 256

Qy 121 KVISHPNDSKTKNNDIAIMKLOKPLTFNDLVKPYCLPMPGMLQPEQLCWSIGMGATEE 180
 Db 257 NILHNHNYKSAHTENDIALVRLNSVTFMDIHSCVCLPATONIPPGSTAYVYTGAGQAEY 316

Qy 181 KAKTSVFNAAKVLITETQRCNSRYVDNLIITAMLCAGPLGNDSCCGSGGKXLY-TS 239
 Db 317 AGHTVPELRQGVRIISNDVCMAPHSTNGAILSGMLCAGVPQSGVDACGDSGGPLVQED 376

Qy 240 KNNIMWLLIGDTSWGGCGAKAYRPGVYGNVWFETDWTYRQ 278
 Db 377 SRRLWFIVGIWSGDQCGLPDKPGVTRVATVADWTIRQ 415

Search completed: August 1, 2003, 18:11:58
 Job time : 13 secs